4

REPORT RESUMES

ED 003 321 24

PHONEME-GRAPHEME CORRESPONDENCES AS CUES TO SPELLING IMPROVEMENT.

BY- HANNA, PAUL R. AND OTHERS
STANFORD UNIV., CALIF., SCHOOL OF EDUCATION
REPORT NUMBER CRP-1991 PUB DATE 65
EDRS PRICE MF-$0.36 HC-$9.24 231P.

DESCRIPTORS- *PHONETIC ANALYSIS, *PHONETICS, *PHONOLOGY, *ORTHOGRAPHIC SYMBOLS, SPELLING, *ENGLISH, LINGUISTICS, STANFORD, CALIFORNIA, COMIT

AN ATTEMPT WAS MADE TO REPEAT, WITH REFINEMENTS, THE RESEARCH DESIGN OF THE HANNA-MOORE STUDY (1953) CONCERNING PHONEME-GRAPHEME RELATIONSHIPS. OBJECTIVES OF THIS STUDY WERE--(1) TO EXAMINE THE SOUND-TO-LETTER CHARACTERISTICS OF AN EXTENDED LIST OF 10,000 WORDS, (2) TO DISCOVER TO WHAT DEGREE THE SAME 80 PERCENT CRITERION OF PHONEME-TO-GRAPHEME OCCURRENCE IS TRUE THROUGHOUT THE AMERICAN-ENGLISH LANGUAGE, (3) TO ESTABLISH AN INDEX OF DIFFICULTY FOR EACH WORD LIST, (4) TO TEST RELIABILITY OF CERTAIN ORTHOGRAPHIC PRINCIPLES, AND (5) TO STUDY THE RELATIONSHIP BETWEEN THE THEORETICAL DIFFICULTY OF SPELLING FROM SOUND-TO-LETTER AND THE EMPIRICAL EVIDENCE OF PUPIL SPELLING PERFORMANCE. AN INDEX OF SPELLING DIFFICULTY OF EACH PHONEME FOR 10,000 WORDS WAS ESTABLISHED. COMIT PROGRAMING PROVIDED (1) LISTS OF WORDS FOR EACH PHONEME, (2) AN INDEX OF DIFFICULTY, AND (3) INFORMATION ON RELIABILITY. THE STUDY WAS CONDUCTED IN TWO PHASES. THIS REPORT IS FOR PHASE ONE ONLY. DATA FROM PHASE ONE SUGGESTS THAT THE AMERICAN-ENGLISH ORTHOGRAPHY MAY BE DESCRIBED IN TERMS OF PHONOLOGICAL STRUCTURE OF THE SPOKEN LANGUAGE. SEVERAL INFERENCES ARE MADE. (JC)

USOE Research Project No. 1991
Stanford University
Paul R. Hanna Director

Text Copy 2 Pages 1-215

[Appendices A, E, f, and I are bound
in 4 separate volumes.]

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

ED 003321

PHONEME-GRAPHEME CORRESPONDENCES
AS CUES TO SPELLING IMPROVEMENT

COOPERATIVE RESEARCH PROJECT NO. 1991

PHONEME-GRAPHEME CORRESPONDENCES

AS CUES TO SPELLING IMPROVEMENT

Cooperative Research Project No. 1991
Stanford University

Paul R. Hanna
Project Director
Professor of Education
Stanford University

Jean S. Hanna
Stanford, California

Richard E. Hodges
Project Research Associate
Assistant Professor of Education
University of Chicago

Edwin H. Rudorf, Jr.
Project Research Associate
Assistant Professor of Education
University of Delaware

1965

The research reported herein was supported by the
Cooperative Research Program of the Office of Education,
U.S. Department of Health, Education, and Welfare.

I

FOREWORD

The authors of this research report were tempted to insert the term Linguistics into the title of this volume in order to capitalize on the mounting interest among educators in reforms of the language arts programs in schools--a movement currently being generated by analyses of the nature and structure of our American-English language. The temptation was resisted with the logic that in this research we are not behaving strictly as linguists; rather we are making a statistical analysis of the degree and the characteristics of the correspondences that exist between the way we speak our language (using the phoneme as the unit of sound) and the way we write our language (using the grapheme as the unit of encoding). Our ultimate purpose is to improve the school spelling program. Therefore we title our report Phoneme-Grapheme Correspondences as Cues to Spelling Improvement.

The origin of this research dates back to 1949 when we posed certain questions regarding the correspondences and their consistency in spelling. James T. Moore, Jr. accepted the challenge as a doctoral dissertation research and in 1951 Hanna and Moore produced evidence that 80 percent of the more than 12,000 phonemes in a 3,000 word list are spelled with predictable regularity. This research was accepted by some spelling programers and in many places changed the approach to school spelling; the research was questioned by others on a variety of grounds. These questions were not answerable in the Hanna-Moore research and led naturally to a desire to delve more deeply into the nature of the phoneme-grapheme correspondences.

It was not until the Cooperative Research Bureau contracted in January, 1963, with Stanford University that finances were available for a massive attack on the problems. During the intervening 24 months of the contract we have studied the sub-problems with computer programs that have produced over 10,000 printout pages of lists, analyses, and statistics regarding the subtle relationships of the correspondences that could not be handled by eye and hand in the previous research effort of a decade earlier.

In these two years we have by no means exhausted the probing. We have answered certain questions rather firmly. There is evidence that the encoding of our aural-oral language is accomplished through the alphabetical principles. We know much more than we did about the degree to which phonological analyses will provide the cues to standard American-English spelling. We also know what percentage of the correspondences must be rooted in morphological and/or syntactical analyses rather than in the sound-to-letter cues alone. We know something about the relative difficulty of spelling any word in the extended list of over 17,000 words. And we know something about the agreement of the theoretical difficulty of spelling as determined by the consistency with which the alphabetical principle is followed and the empirical evidence of spelling difficulty as shown by available research of pupils' spelling performances.

We also know what we must continue to probe. We need to research the problem of dialects and the role they do or do not play in distorting the phonological cues to spelling. We need to refine our knowledge of the morphological and syntactical cues. We need to work out a spelling curriculum and appropriate teaching-learning techniques that will be based on these newer cues. Then we need to give a massive field testing to the new curriculum in order to once again improve spelling competence.

The authors are deeply indebted to many persons for assistance rendered in the execution of this research project. Some of our colleagues have served on dissertation committees; others have served as general consultants; still others have rendered specific assistance in programing, editing, typing, administrative support, etc. We acknowledge our appreciation to each of these and at the same time absolve each from responsibility for errors that occurred in this report. Special thanks are due: Virgil Anderson, Professor of Speech Pathology, Stanford University; Richard G. Axt, Research Administrator (1962), Stanford University; Sidney Bergquist, Research Assistant, Project No. 1991, School of Education, Stanford University; John B. Carroll, Roy E. Larson Professor of Educational Psychology, Graduate School of Education, Harvard University; Edgar Dale, Professor of Education, Ohio State University; Godfrey Dewey, Vice-President, Lake Placid Club Education Foundation; James Dolby, Lockheed Corporation; Joseph H. Greenberg, Professor of Anthropology, Stanford University; Phyllis Groll, Programer (1962), Project No. 1991, Stanford University; Doris V. Gunderson, Research Coordinator, Cooperative Research Branch, U.S. Office of Education; Lois M. Hodges, Secretary to Project No. 1991, Stanford University; Ernest Horn, Professor of Education, College of Education, State University of Iowa; Thomas Horn, Professor of Education, School of Education, University of Texas; Francis A. J. Ianni, Director, Division of Education Research, U.S. Office of Education; Wm. J. Iverson, Professor of Education, School of Education, Stanford University; Alphonse G. Juilland, Professor of French and Chairman of Committee on Linguistics, Stanford University; Robt. B. Kennedy, Contract Administrator (1964), Stanford University; Harry Levin, Professor, Cornell University; Ian MacNeill, Research Assistant, Project No. 1991, Stanford University; Frederick J. McDonald, Associate Professor of Education and Psychology, Stanford University; Bernice McTaggart, Secretary, School of Education, Stanford University; Marcella Miller, Editorial Consultant, Project No. 1991; James T. Moore, Jr., Professor of Education, University of Kentucky; Margaret Moore, University of Kentucky; Robert L. Politzer, Professor of Education and Romance Linguistics, Stanford University; I. James Quillen, Dean, School of Education, Stanford University; Alice Scates, Research Coordinator, Cooperative Research Branch, U.S. Office of Education; Herbert Solomon, Professor of Statistics, Stanford University; John Steffani, Programer, Control Data Corporation; Ruth G. Strickland, Research Professor of Education, Indiana University; Edith C. Trager, Associate Professor of English, San Jose State College; Jeanie Treichel, Secretary to Project No. 1991; Ruth Weir, Associate Professor of Linguistics, Stanford University; Virgil K. Whitaker, Dean, Graduate Division, Stanford University; Axel Wijk, English Institute, Stockholm University; Richard L. Venezky, Consultant to Project No. 1991 on Linguistic and Computer Technology; Fred Zeitler, Programer, Control Data Corporation.

PAUL R. HANNA
Project Director and
Jacks Professor of Child Education
School of Education
Stanford University

December 31, 1964

TABLE OF CONTENTS

Page

FOREWORD ii

LIST OF TABLES ix

LIST OF ILLUSTRATIONS xi

Chapter

1 INTRODUCTION 1

The Structure of Language 3
The Structure of American-English 6
Writing Systems 9
The American-English Writing System 11

2 THE SPELLING PROBLEM: TRADITIONAL AND LINGUISTIC APPROACHES 14

Traditional Understandings of the Spelling Problem . 14
Linguistic Implications for the Problem of Spelling. 17

3 SOME PRELIMINARY DECISIONS 22

The Selection of the Word List 23
The Selection of a Dictionary and Phonemic Code . . 28
The Selection of Graphemic Options 32
The Research Design 35
The Preparation of the Raw Data 37
The Preparation of the Computer Program 42
The Regrouping of the Phonemic Code 49

4 PHASE I: AN ANALYSIS OF THE PHONOLOGICAL STRUCTURE OF AMERICAN-ENGLISH ORTHOGRAPHY 53

Graphemic Options Representing the Phonemes of American-English 53
The Regrouping of Vowel Phonemes 71
Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences 73
The Distribution of Phonemes in American-English . . 83
The Number of Graphemic Options Required to Bring Phonemes to an 80% Criterion 87
Phoneme-Grapheme Correspondences by Position in Syllables 92

TABLE OF CONTENTS--*Continued*

Chapter 4 PHASE I--*Continued*

Chapter Page

- Phoneme-Grapheme Correspondences by Position in Stressed and Unstressed Syllables . . . 111
- The Distribution of Phoneme-Grapheme Correspondences in American-English . . . 141

5 SOME INFERENCES FROM PHASE I . . . 188

- Introduction . . . 188
- Phoneme-Grapheme Correspondences . . . 197
- The Effects of Position in Syllables on Phoneme-Grapheme Correspondences . . . 198
- The Effects of Syllabic Stress on Phoneme-Grapheme Correspondence . . . 200
- Internal Constraints . . . 203

6 A DEFINITIONAL MODEL FOR THE ANALYSIS OF AMERICAN-ENGLISH ORTHOGRAPHY . . . 206

7 PHASE II: THE DEVELOPMENT OF AN ALGORITHM FOR AMERICAN-ENGLISH SPELLING . . . 216

- Analysis Sheets . . . 222
- An Algorithm for American-English Spelling . . . 228
- Known Error in the Algorithm . . . 243
- The Effectiveness of the Algorithm: Computer Results 243
- Graphemic Options . . . 252

8 CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS FROM PHASE II . . . 253

- Conclusions . . . 253
- Recommendations . . . 257

9 THE STUDY IN PERSPECTIVE . . . 261

- Cognitive Processes and their Relationships to Spelling Instruction . . . 261
- Neurophysiology and Its Relationships to Spelling Instruction . . . 262
- Psychology and its Relationships to Spelling Instruction . . . 264
- Some Suggestions for the Spelling Course of Study . 267
- Selecting Words for Spelling Instruction . . . 269
- Constructing Teaching-Learning Strategies . . . 270
- Summary . . . 271

TABLE OF CONTENTS--Continued

Chapter Page

BIBLIOGRAPHY . . . .

APPENDIX . . . .

A. Alphabetized Listing of 17,310 Words Appearing in Corpus.

B. Frequency and Percentage Tabulations of Phoneme Occurrences in Corpus . . . .

C. Listing of Words Containing Ambiguous Spellings . . . . .

D. Listing of True Error Words, Classed by Type of Error . .

E. Vowel Phoneme Word Listing . . . .

F. Consonant Phoneme Word Listing . . . .

G. The Algorithm . . . .

H. Words Spelled Correctly by the Algorithm . . . .

I. Error Listing from Algorithm . . . .

J. Glossary . . . .

LIST OF TABLES

Table | Page
1. Summary of Results from James T. Moore Study: Frequency and Percentage of Regular Spellings by Phoneme Classifications 18

2. Phonemic Analysis Code: 62 Phoneme Classification . . . 39

3. Phonemic Analysis Code: 52 Phoneme Classification . . . 50

4. Graphemic Options Representing Phonemes by Position in Syllables: 22 Vowel Classification 54

5. Graphemic Options Representing Phonemes by Position in Syllables: 30 Consonant Classification 61

6. Additional Occurrences of Certain Consonant Phonemes when < L > is Changed to < LE > as a Graphemic Option for /Ll/ 74

7. Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences in American-English: 22 Vowel Classification 77

8. Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences in American-English: 30 Consonant Classification 80

9. Frequency and Percentage Tabulations of Distribution of Phonemes in American-English: 22 Vowel Classification 84

10. Frequency and Percentage Tabulations of Distribution of Phonemes in American-English: 30 Consonants Classification 85

11. Number of Graphemic Options Required to Increase Selected Phonemes Beyond 80% Criterion: 22 Vowel Classification 88

12. Number of Graphemic Options Required to Increase Selected Phonemes Beyond 80% Criterion: 30 Consonant Classification 90

LIST OF TABLES--Continued

Table | Page

13. Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences in American-English by Position in Syllables: 22 Vowel Classification 94

14. Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences in American-English by Position in Syllables: 30 Consonant Classification 105

15. Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences in American-English by Position in Stressed and Unstressed Syllables: 22 Vowel Classification 114

16. Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences in American-English by Position in Stressed and Unstressed Syllables: 30 Consonant Classification 130

17. Frequency and Percentage Tabulations of the Distribution of Phoneme-Grapheme Correspondences in American-English by Position in Syllables: 22 Vowel Classification . . 144

18. Frequency and Percentage Tabulations of the Distribution of Phoneme-Grapheme Correspondences in American-English by Position in Syllables: 30 Consonant Classification. 153

19. Frequency and Percentage Tabulations of the Distribution of Phoneme-Grapheme Correspondences in American-English by Position in Stressed and Unstressed Syllables: 22 Vowel Classification 162

20. Frequency and Percentage Tabulations of the Distribution of Phoneme-Grapheme Correspondences in American-English by Position in Stressed and Unstressed Syllables: 30 Consonant Classification 176

21. Percentages of Approximation to the Alphabetic Principle of Phoneme Classifications at Each Level of the Phonological Structure Underlying the Orthography . . . 196

22. Summary of the Power of the Algorithm to Spell Words Correctly 244

23. Summary of Results of the Algorithm 246

24. Analysis of Phonological Predictability of Graphemic Options with Distributional Factors Considered 251

LIST OF ILLUSTRATIONS

Figure Page

1. Sample Analysis Sheet: Vowel Phonemes 223

2. Sample Analysis Sheet: Consonant Phonemes 224

3. Coding Format for Algorithm 226

CHAPTER 1

INTRODUCTION

Spelling instruction in American elementary schools traditionally has been predicated on the assumption that there is an inconsistent relationship between the sounds (phonemes) of American English speech and symbols (graphemes) which represent these sounds in writing. Accordingly, many authorities on the teaching of spelling have maintained that each spelling word must be considered a separate learning act. A variety of instructional practices, including the use of mnemonic aids and the teaching of word analysis skills, have dominated a large part of the traditional spelling program.

Current linguistic knowledge suggests, however, a considerably different view of the nature of American-English spelling. An analysis of the nature of the orthography clearly reveals that this writing system is essentially alphabetic in structure. Large numbers of words contain highly consistent phoneme-grapheme relationships (Bloomfield, 1933).

During the 1949-51 academic years basic research into this linguistic observation was conducted at Stanford University by James T. Moore, Jr., directed by Paul R. Hanna. This research demonstrated that 80% of the 12,546 phonemes comprising the 3000 most frequently used American-English words are spelled consistently, while only 20% of the phonemes present unusual difficulty (Moore, 1951). In emphasizing the possibility of using linguistic principles to reinforce other types of learning, the Hanna-Moore study has influenced the current approach to the subject of elementary school spelling. Many schools have designed new programs utilizing a sound-to-letter approach in which the pupil learns that certain specific graphemes are used to write the phonemes in a word he says and hears. The general teaching strategy suggests that the consistently spelled phonemes are introduced first in the developmental program; then those phonemes are taught that have optional spellings. The pupil is helped systematically to observe the consistency of sound and letter and inductively to arrive at basic principles for the orthography of the American-English language (Hanna and Hanna, 1959; Hanna and Moore, 1953).

Some spelling authorities, while acknowledging that the orthography is basically alphabetic, have contended that the research conducted to analyze this structure has been poorly designed. A more refined analysis of additional numbers of American-English words, they suggest, would show clearly that consistent phoneme-grapheme correspondences are found primarily in only the most frequently used words. Analyses of less frequently used words, including derived, inflected, and borrowed forms, would indicate the essentially inconsistent nature of the American-English orthography. These authorities contend traditional spelling techniques still are necessary for the vast majority of words which the pupil must learn to master if he is to be a proficient speller.

It was the major purpose of the research reported here to analyze phoneme-grapheme correspondences in a 10,000 or larger word list in order to clarify the issues which exist between these two schools of thought regarding the spelling curriculum and also to determine whether such a depth analysis of the language itself would yield other cues useful for the improvement of spelling.

The Structure of Language

Language is the principal means of human communication. Of the several thousand known languages of the world, each may be described as an arbitrarily selected communications system since no two languages utilize identical sound features, yet organize their essential sound features into systematic and coherent patterns or structures. Classifying the different structures of the many language systems in an effort to make explicit what language users know implicitly is one of the central undertakings of linguistic science.

Languages are systems of verbal habits which children learn by imitating older children and adult models. The language learner first identifies the basic phonemic stock of the language system and then learns to arrange these sound features into appropriate patterns which produce words that are included in the lexicon. In spite of the fact that different language communities select different sets of phonemes to carry on oral communication, each oral language system can be classified in a hierarchical arrangement which is essentially similar for all languages [illegible] (Spang-Hanssen, 1959, p. 28).

These structural components of oral language include: (1) a phonetic reservoir from which a phonemic code is selected, (2) the phonemic code itself, (3) the arrangement of phonemes into speech units which minimally express meaning (i.e., the morphology of the language), and (4) the arrangement of morphemes into syntactic patterns which (5) convey meanings in terms of the conceptual system of a language community.

In terms of the historical development of language systems, as well as in the order in which spoken and written communication is learned, the oral language code is primal to the written code (Hall, 1961, p. 1).

A written language is, in sum, a visual surrogate of oral language, although the degree to which different language communities develop writing systems in close approximation to oral communication varies widely. In order more thoroughly to understand what is meant by phoneme-grapheme correspondences, as referred to in this report, it is necessary first to describe briefly the structural components of language systems generally.

The phonetic reservoir. The total stock of undifferentiated sounds which human beings are physiologically capable of making forms the phonetic base out of which languages are constructed. The classification of these undifferentiated sounds in terms of their acoustical and articulatory properties lies in the domain of a particular branch of linguistics--phonetics. The phonetician describes how speech sounds are formed and does not concern himself necessarily with the ways in which phonemes are selected and arranged into meaningful utterances.

The phonemic base. From this phonetic reservoir are selected those distinctive sounds which form the phonemic base of a particular language system. These are, in turn, arranged in various patterns to produce meaningful utterances. The number of minimally distinctive sound features, or phonemes, which have been selected by known language communities ranges from approximately 13 for Hawaiian to approximately 75 for a language community of the Northern Caucasus (Hockett, 1958, p. 93). It is estimated that approximately 45 phonemes seem to be the average number of different phonemes used in the known languages of the world.

Although some linguists consider phonemes to be abstractions, or averages of the total number of ways in which certain distinctive sound features may be pronounced, they cannot be considered pure abstractions since their acoustical and articulatory properties are quite real (Gleason, 1951, p. 19). In any case, phonemes are the basic constituents of oral language systems.

The morphological base. The arrangement of phonemes into minimal units of meaning constitutes the morphological base of oral language. Within the limitations which a language system places upon acceptable phonemic combinations, many phonemic substitutions may occur in the morphemes used in the lexical stock. For example, bit will become, by substituting phonemes in various positions in this word: pit, bet, or bin. Certain environmental restrictions do occur, however, since some phonemes cannot be legitimately used in certain positions in some phonemic sequences. Thus, /NG/[1] is not used in initial position in American-English words.

The syntactic and grammatical base. The arrangement of word forms into larger structural units (e.g., noun and verb groups, prepositional groups, clauses, and sentences) comprises the syntax of language systems. Each language has implicit processes and principles by which syntactic patterns are constructed. In turn, the total structure of a language system constitutes the grammar of that system.

Once again, using American-English as an example, consider the different meanings which are signaled in the following utterances when the word little is used in different syntactic patterns. The little boy realized his mistake. The boy little realized his mistake. The boy realized his little mistake.

---

[1] Throughout this report, an intended phoneme or bound morpheme will be indicated by / /, and an intended grapheme will be indicated by < >. For example, the first sound heard in the word pin is the phoneme /P/. The grapheme which represents this phoneme is < P >.

The semantic base. Ultimately, the purpose of any language system is to permit intelligible human intercommunication. Language is the code permitting intended meanings to be conveyed from one member of a speech community to others in that community by means other than direct experience. Language is a surrogate of human experience, just as writing systems are surrogates of spoken language. The phonemes, morphemes, and syntax of a language system are the instruments by which this human intercommunication is made possible. The meanings conveyed by language are the semantic properties of the system, although more properly, perhaps, a part of the study of semantics is the study of human behavior itself.

One of the more common misunderstandings of language systems is to mistake the words included in these systems for the meanings they convey. The words which form the lexicon of a language community refer to man's conceptualizations of the world in which he lives. Thus, not only are different sets of phonemes used in different language systems, but these phonemes are arranged into different patterns or words to refer to these conceptualizations. In addition, the conceptualizations themselves are contingent upon the experiences which are available to members of respective cultures. These factors suggest some of the complexities involved in accurately translating the utterances of one language community into those of another.

The Structure of American-English

This study is concerned with the American-English language. As a major dialect of the English-speaking language system, American-English is differentiated from other dialects by the allophonic variations of the phonemes of the English language generally. Thus, American-English is in reality one dialect of the general English-speaking community. It shall be defined in this study, however, as a language system, since there are numerous regional dialect differences within the system, and since American-English is generally considered by its users to be a language system quite distinct from other dialects of the English-speaking world.

As in other languages, a classification of which phonemes are ordered into syntactic patterns to permit mutual intelligibility of spoken intercommunication also describes the structure of American-English.

The phonemic base of American-English. Approximately 46 distinctive sound features are selected by many linguists from the phonetic reservoir to form the phonemic base of American-English. Of these, 24 phonemes are consonantal, nine phonemes are vocalic, three phonemes are semi-vocalic, and ten are the suprasegmental phonemes of pitch, stress, and juncture (Gleason, 1961, p. 50).

Pronounceable units of sound are typically formed around a vowel nucleus. To these vocalics are added consonant phonemes which, in combination, provide the syllables of the American-English oral language system. Syllables, in turn, are then arranged in various combinations to produce the lexical stock of American-English.

The morphological base of American-English. American-English is primarily an analytic language system, combining root forms with a number of different affixes which add to and alter the minimal meanings of the morphological stock. Rules of affixation are comparatively consistent, although some irregular forms do occur. The morphemes /-s/,

/-z/, and /-ez/, for example, are the dominant plural affixes used in American-English, while /-ren/ in children is so irregular as to be unique. In general, however, the morphological base of American-English is relatively stable, having comparatively few irregular forms.

The syntactic and grammatical base of American-English. The lexicon of American-English is arranged into acceptable syntactic patterns for the purposes of meaningful oral communication. This grammatical structure traditionally has been described in terms of parts of speech and in terms of the meanings to which these speech parts refer. Nouns, for example, commonly are said to be the part of speech which refers to "persons, places, and things."

To the linguist, however, American-English syntax is more clearly described in terms of how these speech forms function in contributing to the total meaning of an utterance. Thus, the word bear would be classified in traditional terminology as a noun or verb, depending on the meaning bear is intended to convey. In linguistic terminology, however, bear is a member of the form class "noun" in the bear ran down the road, and is a member of the form class "verb" in the man will bear arms against the common enemy, thus emphasizing the functions these words perform rather than their meanings.

These rules for syntactic arrangement are learned implicitly during the course of an individual's language development, and these grammatical constructions are, in turn, judged acceptable or unacceptable in usage upon the basis of social class standards. In short, the rightness of grammar (its grammaticalness) is not universal in American-English.

The semantic base of American-English. As with all language systems, the ultimate purpose of combining phonemes into morphemes and morphemes into syntactic patterns in American-English is to communicate intelligibly. In American-English, however, a special group of words, homonyms, create certain difficulties in both oral and written communication. In spoken language, homonyms, or more accurately, homophones, are words having identical pronunciations but different meanings. The meanings of homophones are ordinarily differentiated by the availability of syntactic cues. In some instances, homonyms are spelled differently, thus providing additional cues to their meanings in written language. However, many homonyms also have identical spellings so that these homographs can only be differentiated semantically upon the basis of syntactic cues. For example, the sentences: the tree is bare of leaves; the tree will bear leaves; and the bear climbed the tree illustrate the kinds of semantic confusion which might arise were not syntactic and graphic cues available to the listener or reader in attributing the appropriate meanings to bear and bare.

Writing Systems

Written language, as a visual surrogate of speech, is predicated upon the structural characteristics of spoken language. The symbols used in writing, regardless of the language system, are graphic representations of the phonology, morphology, and syntax of speech (Francis, 1958, p. 438).

The selection of graphic symbols, like the selection of phonemes, is largely an arbitrary matter. These graphic representations may be used to visualize not only phonemes, as in American-English, but also syllables and, in some cases, words. Those language systems using a single graphic symbol to represent entire words are logographic in structure. No known language community uses logographic writing entirely, although Chinese has often been described in this way. More accurately, the Chinese writing system is a combination of logographic and morphographic writing, the latter form using a single graphic symbol to represent the morphemes of the spoken system. In American-English, which uses very few logographic and morphographic referents in the writing system, the ampersand (&) typifies logographic writing, while the apostrophe in boys' is morphographic and signals the plural possessive reference of the bound morpheme /-z/.

Some language communities use single graphic symbols to represent the syllables of speech and are thus described as syllabic writing systems. The syllabaries of Japanese are representative of this form of writing, although, like Chinese, some morphographic writing is combined with syllabic writing as well. In American-English I C LN (I see Ellen) would, if used, typify a combination of logographic and syllabic writing.

Perhaps the dominant writing system used among the world's language communities is alphabetic, or phonemic writing, in which each graphic symbol represents a particular phoneme of speech. Alphabetic writing is considered the most versatile of all writing systems since the number of different graphemic combinations possible is nearly as great as the number of phonemic combinations permitted in a particular language community. Certain restrictions occur, however, to deter the most efficient use of this written code.

Ideally, alphabetic writing would always use one graphemic symbol to represent one phoneme. Given an oral language system comprised of 45 phonemes, a similar number of graphemic symbols would be employed in writing. No writing system, however, is quite so consistent, since suprasegmental phonemes often do not have graphic symbols representing them. Some language communities, such as Finnish, Italian, and Hawaiian, do come very close to purely alphabetic writing systems in which a single phoneme is always matched by a single grapheme (Gelb, 1952). For speakers of these languages, learning to spell is a relatively simple matter. If one knows which graphemes always represent the phonemes of speech, the spelling of a particular word is largely a matter of writing the graphemes in the same sequence in which the phonemes are arranged. For such languages, the alphabetic principle, to all intents and purposes, is fully achieved.

The American-English Writing System

Speakers of American-English do not enjoy the confidence that they can always transcribe accurately the phonemes of speech into the graphemes of writing. Although basically an alphabetic writing system, the American-English orthography is handicapped by having only 26 single graphemic symbols (the letters of the alphabet) to represent approximately 36 segmental phonemes and virtually no graphemic symbols which explicitly are intended to represent the pitch, stress and juncture suprasegmental phonemes. Thus, some graphemic symbols necessarily must represent two or more phonemes, while such punctuation symbols as < ? > denote not only the kind of utterance being made but also indicate the pitch pattern which ends the utterance.

American-English orthography is further complicated by the fact that phonemes are often represented by combinations of graphemes so that graphemic options for phonemes may be comprised of as many as four graphemic symbols. For example, the phoneme /O/ in the word though is represented by the graphemic option < OUGH >. When the orthography is described in this way, it can be seen that the problem of spelling is not merely a matter of assigning an insufficient number of graphemic symbols to phonemes. One of the principal difficulties of American-English orthography is the fact that there are many more graphemic options than there are phonemes.

Furthermore, the American-English lexicon includes many words borrowed from other language systems and thus incorporates their written forms as well. These words often are not based upon the same phoneme-grapheme relationships which underlie the American-English orthography. In addition, the evolution of American-English from general spoken and written English has resulted in the creation of certain departures from the alphabetic principle which do not occur in the parent language system.

Many of the graphemic symbols of contemporary American-English writing no longer represent the same phoneme to which they earlier referred during the period when the orthography was standardized. As a result, such words as of, where < O > represents the vowel sound heard in cup and < F > represents the consonant phoneme /V/, have little relationship to the alphabetic principle underlying the orthography. Such words, especially those that have low frequency of usage in writing, are some of the major reasons that the American-English orthography is described as being "archaic, cumbrous, and ineffective" (Hall, 1961, p. 33).

Yet, in spite of such inconsistencies among phoneme-grapheme relationships in the orthography, most American-English writers learn implicitly to assign appropriate graphemes to the phonemes of spoken words, although with varying degrees of exactness. Since the alphabetic nature of the writing system has not been taught explicitly in most American schools, it is the fortunate and determined pupil who intuitively recognizes and applies the alphabetic principle in his writing.

However, increasing efforts are being made to teach directly the alphabetic principle in spelling programs so that pupils can more efficiently learn the rules of the orthography and apply them in their writing. That the alphabetic structure of large numbers of words has not been identified has been the result of limitations in research design and technological facilities with which to carry out so large a task.

Recognizing that the alphabetic principle can only be approximated in American-English, and that many words--or certain parts of many words--still must be learned by means other than direct application of this principle, the study reported here has sought to analyze as large a corpus of words as was practical, using modern computer technology, and to classify the graphemic representations of these words in terms of how closely they do approach the alphabetic principle according to the positions which phoneme-grapheme correspondences occupy in syllables. The findings of this research, it can be argued, could have a considerable effect on the way the content of spelling curriculums is organized and taught in American classrooms. A review of how this curriculum traditionally has been conceived, and how contemporary linguistic knowledge has relevance for spelling programs, is necessary for an understanding of the rationale underlying the research herein reported.

CHAPTER 2

THE SPELLING PROBLEM: TRADITIONAL AND LINGUISTIC APPROACHES

Traditional Understandings of the Spelling Problem

It has been noted that traditional spelling programs have made little use of the alphabetic principle underlying the American-English orthography. Accordingly, it has seemed necessary to educators to consider each spelling word as a separate learning act since very few words appear to contain consistent phoneme-grapheme relationships.

Spelling programs in elementary schools thus have included many and varied kinds of instructional methodologies. It is true, of course, that traditional instructional practices were developed before the insights of linguistic science became evident to spelling curriculum specialists. Yet, even today, an examination of contemporary literature on spelling instruction indicates that little use is made of linguistic information in elementary school spelling programs (Hanna and Hodges, 1963).

Since the American-English orthography traditionally has been conceived to have little correspondence with the spoken language, spelling words must be selected and grouped on some basis other than the alphabetic principle underlying the orthography. Many rationales for word selection and grouping have been put forward by spelling authorities in an effort to help pupils learn their spelling words.

One of the more typical approaches to the selection of words for spelling instruction has been to determine which words children and adults will most likely need when they write (Fitzgerald, 1951). A great deal of research has been conducted to determine which words of the American-English lexicon are most frequently used and the order in which these words should be presented in spelling programs (Dale, 1960; Greene, 1954; Horn, 1926; Ringland, 1945; Thorndike and Lorge, 1944).

Typically, weekly spelling lessons have been built around a list of about 10 to 20 words that may be classified in different ways:

1) Words grouped at random (e.g., _tree_, _fine_, _sick_)
2) Words grouped according to visual similarities (e.g., _nation_, _function_, _invitation_)
3) Words grouped about a spelling rule or generalization (e.g., words that end in _y_, change the _y_ to _i_ before adding suffixes, or the _es_ of the plural form.
4) Words grouped into meaningful association around a typical child interest (playing with dolls) or a curricular topic (Colonial Life)
5) Words grouped in phonemic families (e.g., long _a_ sound as in _make_, _made_ spelled _a_ and a final silent _e_ (Hanna and Hanna, 1959, pp. 12-13).

Once a particular rationale has been chosen by which words can be classified for spelling instruction, the next step is to determine which words should be selected that best illustrate the chosen rationale. Various criteria have been developed, but, in general, the traditional word list would consider:

1) the relative _frequency_ with which the word is used,
2) the _commonness_ of the word's use,
3) the _spread_ of the word's use in different types of writing,
4) the degree of _cruciality_ possessed by the word as evidenced by the severity of penalty attached to its misspelling,
5) the probable _permanency_ of the word's use,
6) the _quality_ of the writing in which the word was used, and
7) the _persistent difficulty_ of the word (McKee, 1944, pp. 154-55).

Traditional spelling programs likewise have placed a heavy emphasis upon instructional methodology. Curriculum specialists have recognized the fact that the words selected for study according to criteria such as those noted above have no inherent structural relationship. Thus, the processes of instruction must be the most effective possible. Some authorities have suggested that words be studied in terms of their syllabic structure and the root form from which such words are built (Osburn, 1954). Others have recommended an instructional methodology built around pre-testing and post-testing of words in the weekly spelling list (Fitzgerald, 1952) and the conscientious correcting of spelling errors made on these tests (Horn, 1947).

Another instructional rationale has attempted to classify children's spelling errors in writing and to develop instructional processes which stress the parts of words which apparently are the most difficult for children to spell (Gates, 1937).

In sum, a fundamental assumption regarding the teaching and learning of spelling, from a traditional point of view, has been a heavy reliance upon visual learning (Gilbert and Gilbert, 1942).

Visual learning necessarily assumes that one learns to spell after having learned to read, since the writing system is a graphic representation of oral language, imperfect though this system might be. Spelling, therefore, is essentially a special problem in reading wherein the child must learn to pronounce the letters of written words and remember which letters "make" the sounds of these words. A need for careful pronunciation of the "letter sounds" is a necessary part of this instructional approach (McKee, 1944).

That the letters do not have sounds (e.g., the five vowel "sounds" of the letters < A >, < E >, < I >, < O >, < U >, and sometimes < Y >), but that instead they represent in varying degrees of correspondence the sounds of the spoken language, has been one of the most cogent revelations which linguistic science has made evident.

Linguistic Implications for the Problem of Spelling

Most linguistic studies of the structures of spoken and written languages have been in the area of pure research, although some investigators have had subsidiary motives. Godfrey Dewey (1923), in an early study of the frequency of speech sounds in English, was interested in the possible implications of his work for spelling reform; French, Carter, and Koenig (1930) investigated the frequency and distribution of phonemes in American-English as a possible source of data relevant to the improvement of telephone communications.

Leonard Bloomfield's attempts to apply directly the evidence from linguistic science to the teaching of language skills to American children has had a fundamental effect upon spelling instruction. In a book directed primarily to students and scholars of linguistic science, Bloomfield (1933) explicitly noted the basically alphabetic nature of American-English writing and the possible value of expressly designing reading and spelling instruction around the alphabetical principle.

Some years later, Bloomfield, in an article addressed to teachers of English (1942) suggested the pedagogical values of devising reading programs based upon linguistic criteria. He noted that:

1) words in print are representations of words in speech;
2) initial instruction in reading be based upon learning those words which consistently represented in graphemic form the phonemes of speech;
3) and that once mastery of the regular forms of English writing had been attained, that increasingly irregular forms be gradually introduced in the children's reading and writing vocabularies.

In sum, Bloomfield advocated a rather radical departure from traditional attitudes concerning the teaching of reading and spelling. He stressed the need to teach pupils skills in associating spoken language to its written form, rather than the reverse process of writing into speech. The actual number of words which would fall into groups based on the alphabetic principle, however, remained to be tabulated.

The first basic research undertaken to analyze the American-English language in terms of Bloomfield's hypothesis was initiated as a doctoral dissertation by James T. Moore, Jr. under the direction of Professor Paul R. Hanna of Stanford University (1951). Examining Bloomfield's hypothesis, Moore tabulated the 3000 most frequently used words in children's oral and written vocabularies in terms of the frequency and percentage of phoneme-grapheme correspondences, using a phonemic classification which singled out vowels, single consonants, consonant blends, suffixes, and final blends. He concluded that an overall 80% "regularity" of phoneme-grapheme correspondences was attained in these 3000 words. Table 1 summarizes these findings. The results of Moore's

TABLE 1

SUMMARY OF RESULTS FROM JAMES T. MOORE STUDY:[a] FREQUENCY AND PERCENTAGE OF REGULAR SPELLINGS BY PHONEME CLASSIFICATIONS

Group of Phonemes	Total Frequency of Regular Spellings	% of Regular Spellings	Gross Frequency for Group	No. of Phonemes
Vowels	3,059	65.39	4,678	25
Single Consonants	5,242	91.26	5,744	17
Consonant Blends	1,312	92.19	1,423	55
Suffixes and Final Blends	427	60.91	701	18
For all Speech Sounds	10,040	80.02	12,546	

[a] Adapted from Moore (1951).

research clearly suggested the feasibility of designing elementary school spelling curriculums based upon an explication of the alphabetic principle. Both Hanna and Moore proposed that:

> 1) Some time must be set aside during the school day for a concentrated attack on the business of learning to translate sounds into written symbols.
> 2) Spelling must be integrated with other subjects in the curriculum in order that there may be an emphasis on meaning and correct usage, and a practical application of the proficiency acquired during the regular spelling period.
> 3) There are definite groups of words and syllables in the English language which belong in certain phonic categories. The child should learn such group patterns inductively. He should develop a sense of the probable letter or letters to be used to represent the speech sounds as they occur in words belonging to such group patterns.
> 4) The beginner must proceed slowly from the simple phonic pattern used to write simple words to those groups of words which are complex in phonic structure.
> 5) The relatively few English words or parts of words which follow a rarely occurring phonic pattern (the irregulars) must be individually memorized (1953, p. 337).

These conclusions, however, were challenged by many spelling authorities. Ernest Horn, one of the more scholarly spokesmen, maintained that less frequently used words beyond the 3000 word sample analyzed in the Hanna-Moore study might show less consistent phoneme-grapheme regularities. To substantiate his contention, Horn analyzed 10,000 words, using a different research design intended to overcome some of the methodological weaknesses he cited in the Hanna-Moore study. His conclusions were sharply disparate with the findings of Hanna and Moore (1957).

In a subsequent article in the Encyclopedia of Educational Research (1960), Horn expanded his position regarding the alphabetic nature of American-English orthography. He noted:

1. More than one-third of the words in a standard reference work on the pronunciation of American-English showed more than one accepted pronunciation.

2. Most sounds can be spelled in many ways, one spelling not being sufficient to call it the most "regular" spelling.
3. Over one-half of the words in a conventional dictionary of American-English contain silent letters, and about one-sixth contain double letters when only one letter is actually sounded.
4. Most letters spell many sounds, especially the vowels.
5. Unstressed syllables are especially difficult to spell.

Such admonitions were not meant necessarily to refute attempts to devise spelling curriculums that included, in part, instruction in the alphabetic principle. Rather, Horn suggested that the limited number of words wholly consistent with this principle precluded the development of curriculums which primarily are based upon such linguistic evidence. Horn advocated teaching consistent phoneme-grapheme relationships whenever pertinent, but maintained that, in the main, "each word should be selected on the basis of the extent and nature of its distribution among the various types of writing that have been sampled" (1960, p. 1340).

Subsequent investigations of American-English orthography, principally by interested linguists, have drawn rather different conclusions from Horn's (Garvin, 1963; Hall, 1961; Lloyd and Warfel, 1956; and Porter, 1961). On both empirical and logical grounds, these investigators have re-emphasized the basic validity of the implications of the Hanna-Moore study. Their findings have clearly established that the structure of written American-English is essentially alphabetical.

The degree to which the phonological structure consistently is represented throughout the lexicon, however, had not been investigated in great depth. The need to investigate thoroughly phoneme-grapheme relationships in as large a corpus of American-English words as possible was obviously apparent. The research reported herein has attempted to make this kind of extensive analysis. Devising a research design based upon linguistic principles, and utilizing the powerful research tool of computer technology, over 17,000 different words from the lexicon were analyzed to determine how closely the American-English orthography approximates the alphabetic principle.

CHAPTER 3

SOME PRELIMINARY DECISIONS

The Hanna-Moore research was critized upon several grounds, principally that:

1. Three thousand words are insufficient samples of the American-English language to determine the amount of phonemic-graphemic relationships in the orthography.
2. The separate classification of "silent" letters erroneously suggests a higher degree of phoneme-grapheme consistency than would be the case were such letters assigned to some sound.
3. The classification of consonant blends and the blend of the schwa sound followed by r as graphemic units rather than as separate sounds with separate graphemic representations created an artificially high percentage of phoneme-grapheme consistency (Horn, 1957).

To verify generally the findings of the Hanna-Moore study in a substantially larger number of words and to pursue further linguistic inferences from this study, a much more detailed analysis of the American-English orthography, Project No. 1991, "Phoneme-Grapheme Relationships Basic to Cues for Improvement of Spelling," was initiated at Stanford University in January, 1963, under the sponsorship of the Cooperative Research Branch of the United States Office of Education, Department of Health, Education, and Welfare. The project was directed by Professor Paul R. Hanna and Mrs. Jean S. Hanna with Richard E. Hodges and E. Hugh Rudorf as project research assistants. In consultation with several members of the Stanford community and consultants from other institutions, a linguistically oriented research design with which to study the orthography was developed and applied.[1] Before the actual research could be undertaken, however, several important decisions had first to be made.

The Selection of the Word List

The first problem facing the investigators was the selection of a word list which would be so inclusive that it would be an adequate representation of the total American-English lexicon. In addition, a phonemic code had to be selected, and adapted, by which phoneme-grapheme correspondences in these words could be classified.

Since one of the major criticisms of the Hanna-Moore study had been the limited number of words analyzed, it was decided to go as far beyond the 3000 most frequently used words as time and project finances would permit. Ideally, the total American-English lexicon would furnish incontestable testimony regarding the degree to which the orthography approximates the alphabetic principle. However, no reliable source of the American-English lexicon is available inasmuch as a classification of this population can only be hypothetical. Some words which have only limited frequency of usage tend to become archaic, while new words continually enter the lexicon. As a result, no investigator can be certain that at any given moment a lexical listing accurately includes all words actually used by American-English speakers and writers. Thus, a representative sample of the lexicon is reliable to the extent that it is based upon sound logical assumptions. For this reason, the investigators set the criterion for selecting the number of words to be analyzed in this study as the hypothetical speaking and writing vocabularies of the "average educated American."

---

[1] See Foreword for a list of faculty participants and outside consultants associated with the project.

It is obvious, of course, that the physical chemist has many words in his verbal repertoire which are unique to his area of specialization, words which would not be used by members of other professions. Such words are normally learned while the individual prepares for his profession and while he is engaged in his career. Except for occasional reference to such professions in the literature and in classroom discussion, words peculiar to a particular occupation are not a part of the vocabulary of the majority of American-English speakers and are not usually taught in elementary schools.

Indeed, the legitimacy of estimates of the total number of words the "average educated American" employs in his speaking and writing is open to question. The so-called "common core" vocabulary which has been suggested by many authorities shows a surprising range in the estimates of the size of such vocabularies (Lorge and Chall, 1963).

It was therefore decided, somewhat arbitrarily, that a 20,000 word list would provide as near an estimate of the total lexicon as would be useful in determining phoneme-grapheme relationships in American-English writing. Twenty thousand words (or as near to that figure as possible) would, in the opinion of the investigators, be representative of the vocabulary of the "average educated American."

Selecting a source from which to extract this large number of words was not as obvious as might be assumed. One logical choice was the Thorndike-Lorge Teacher's Word Book of 30,000 Words (1944). But a careful examination of this word source showed several limitations. For one, the basic list had been prepared over twenty years ago and thus does not contain words which since have entered the American-English language.

Concomitantly, the Thorndike-Lorge word list does not account for changes in frequency of usage among the words listed which have occurred since publication of the list.

Another limitation of the Thorndike-Lorge word list was the large number of words included due to the type of literature on which the word counts were based. Words found in Biblical writings, in Shakespearean literature, and in eighteenth and nineteenth century novels are included in the word list and are not generally representative of the speaking and writing vocabularies of contemporary American pupils. Many other kinds of words in this list are extraneous to a "common core" vocabulary of American-English. Nonetheless, in spite of these restrictions, and after considerable searching for other possibly useful word sources, the investigators selected the Thorndike-Lorge Teacher's Word Book of 30,000 Words, Part I, as a reliable classification of the most frequently used words in the lexicon.

Part I of the Thorndike-Lorge list contains 19,440 entries, and it was from this source that the words to be analyzed were taken. Excluded from the basic list were the following entries:

1. Proper names
2. Contracted word forms
3. Hyphenated words
4. Abbreviations
5. Archaic and poetical words
6. Foreign words
7. Trade names
8. Slang and dialectic words
9. Words indicated as "now rare" in standard dictionaries

10. Words for which no pronunciation was provided in the dictionary that subsequently was chosen for this research.

The deletion of these word types reduced the basic list to 15,284 entries. It was thought feasible to supplement this list with additional numbers of words which were either new to the lexicon or had been excluded from the Thorndike-Lorge list due to low frequency of usage. An examination of the "Addenda Section" of the last printing of the Merriam-Webster's New International Dictionary, second edition, unabridged (1957), revealed, however, that no pronunciation was provided for the entries in the Addenda. Since standard entries for both the pronunciations and graphemic representations of words were necessary for the proposed analysis, this source of recently introduced words to the lexicon was eliminated.

It was then decided that the two project research assistants select from the Merriam-Webster's New Collegiate Dictionary, sixth edition (Bethel, 1961), those words which were considered to be relevant to a study of a "common core" vocabulary but had not been available from Part I of the Thorndike-Lorge word list. The selection of such words is, of course, quite subjective. But it was assumed that both Mr. Hodges and Mr. Rudorf would provide checks upon each other's biases in selecting words which would be peculiar to their particular educational and cultural backgrounds.

As a result, Mr. Hodges selected words from the first half of the designated dictionary, while Mr. Rudorf selected words from the second half. Two criteria were established for the word choices made:

1. Derived word forms and root words which already were included in some form in the basic word list would not be selected for adding to the list unless a significant change in their pronunciations was noted.

2. Words representative of areas of professional specialization would be selected if, in the opinion of the researchers, they were sufficiently used in contemporary speech and writing to be considered a part of the common core vocabulary of most educated Americans.

One of the factors the research team wanted to preserve for possible later study was the frequency count number provided for each word in the Thorndike-Lorge list. This list, however, classifies words according to the number of occurrences per million words. Part I classifies 1069 words appearing 100 or more times per million (the "AA" words), 952 words appearing 50-99 times per million (the "A" words), and 17,419 words occurring in decreasing order of word usage from 49 to one occurrence per million words of running text. So fine a discrimination was not considered essential to the research, and so these frequency numbers were regrouped into a less refined frequency code as follows:

Thorndike-Lorge Frequency Code	Regrouped Frequency Code
AA - 100 or more occurrences per million running words	1
A - 99-50 occurrences per million running words	2
49-30 occurrences per million running words	3
29-10 occurrences per million running words	4
9- 1 occurrences per million running words	5

A "tag" code number 6 was then given to the words which were added from the Merriam-Webster's New Collegiate Dictionary. The combined total of 15,284 words selected from the Thorndike-Lorge Teacher's Word Book of 30,000 Words and the 2026 words (including 16 derivations of words already included from the Thorndike-Lorge list) selected from the Merriam-Webster's New Collegiate Dictionary provided a 17,310 word corpus upon which to base the proposed research.

It is interesting to note that only slightly more than 2000 words were added to the basic list. Both investigators have had several years of education past the baccalaureate degree and both have particular interests in studies of the American-English language. In spite of these interests and specialized training in language studies, the comparatively few words which were eventually added to the basic list is indicative of the fact that a 17,000 word speaking and writing vocabulary is quite substantial. Indeed, it can be assumed that, for the purposes of general spoken and written communication, a 17,000 word corpus is closely representative of the total American-English lexicon of an educated United States citizen. It is true that the particular backgrounds of the investigators caused them to select words which other persons might not have chosen. Yet, the surprisingly small number of words which were finally added to the basic list suggests that any differences in the selection of these additional words would be insignificant.

The Selection of a Dictionary and Phonemic Code

No other aspect of this investigation is likely to generate as much discussion, perhaps controversy, as the dictionary chosen by the research team upon which to base an analysis of phoneme-grapheme correspondences. Much effort was expended in order to select a phonemic code which would best compromise the many objections made to contemporary dictionaries and their pronunciation guides.

For reasons which will be set forth, the Merriam-Webster's New Collegiate Dictionary, sixth edition, was selected (Bethel, 1961). It was stated at the outset that this research is based upon linguistic principles. Thus, it seemed logical to select a phonemic code which reflects linguistic knowledge and has general acceptance among linguists and other investigators of spoken and written American-English. Two such sources which are considered by many to meet this standard are: Webster's

Third New International Dictionary of the English Language, unabridged (Gove, 1961), and Kenyon and Knott's A Pronouncing Dictionary of American English (1953).

Recognizing that these are generally acknowledged as standard references for the pronunciation of contemporary American-English, serious consideration was given them as possible sources of phonemic classifications to which the 17,000 word corpus could be compared. A crucial matter arises, however, concerning one of the major purposes of this research. The investigators were interested in eliciting cues to the improvement of spelling, a process which involves the translation of the phonemes of speech into the graphemes which represent them in writing.

Accurate spelling involves providing the correct graphemic option for each of the phonemes in spoken words. Accordingly, a clear pronunciation of these words is essential if the speller is to know which phonemes are to be represented graphemically. A most useful instructional technique in developing such oral-aural cues is to enunciate clearly each spelling word, taking it out of the context of verbal utterances in which it might be used. It is recognized, of course, that such "spelling pronunciation" does some injustice to the way phonemes are normally pronounced and heard in words in running speech. In normal conversation, poor enunciation of individual phonemes is tolerated since the message itself is the prime concern and some redundancy is built into the spoken language. Both Webster's Third New International Dictionary and Kenyon and Knott base their phonemic codes on the pronunciation of phonemes as used in conversations. As Kenyon and Knott have noted,

> Colloquial pronunciation is here treated as the conversational and familiar utterance of cultivated speakers when speaking in the normal contacts of life and concerned with what they are saying, not how they are saying it (1953, p. xvi).

To the descriptive linguist, such a source of American-English phonemic pronunciation is quite useful. But, for the purposes of spelling instruction this form of pronunciation is not the most effective. Much more efficient for an initial introduction to words in daily spelling lessons is the exaggerated pronunciations of phonemes in words said in isolation from context. As Lloyd and Warfel have remarked,

> Spelling is more closely related to words spoken in isolation than to words integrated into running speech. When we pronounce a word by itself, we give it word-stress rather than stresses built into the overall stress-patterns of the utterance, and we give each sound in it a fuller treatment. We consciously pronounce it as we think it ought to be pronounced, often even saying it syllable by syllable.
>
> . . . .
>
> Good spellers have one way of saying words in isolation that does emphasize their spelling and help to recall the standard forms of writing. Knowing which words they have had trouble spelling, they give an arbitrary sound-value to each letter in them (1956, p. 354).

It is this kind of spelling pronunciation that accentuates the use of oral-aural cues in correctly spelling many American-English words.

An exaggerated pronunciation also enhances the possibility of taking full advantage of morphological cues in many spelling words. As an example, the word debate is given the pronunciation spelling of /dəˈbāt/ in Webster's Third New International Dictionary and /dɪˈbet/ in the Kenyon and Knott dictionary. In running speech, the first vowel phoneme thus typically is either the first vowel sound heard in account, or the vowel sound heard pit. For oral-aural cues to spelling debate, however, such pronunciations may be confusing. By pronouncing clearly each syllable in debate, so that the prefix /de-/ is said and heard with the vowel sound heard in be, the pupil is helped to associate the graphemic option < E > with the first vowel sound which it represents.

Both the Webster's Third New International Dictionary and the Kenyon and Knott dictionary use phonemic codes which obscure the spelling pronunciations of words and would, therefore, seem to tend to confuse pupils setting out to learn the American-English orthography. On the other hand, a review of the literature regarding dictionaries and their construction reveals that the Merriam-Webster's New International Dictionary, second edition, unabridged (Neilson, 1957), uses a pronunciation key based upon the speech of public speakers during the time when amplification systems were not yet developed (Eisengon and Ogilvie, 1957). In such instances, running speech had to be precise, even pedantic, lest audiences fail to hear correctly the phonemes of the words which carried the content of the speaker's message.

Granting the validity of the reasoning involved in the selection of a phonemic code for this research, it was decided that the pronunciation key of the Merriam-Webster's New Collegiate Dictionary, sixth edition, identical with its parent dictionary, would provide the kind of general American-English "dialect" most suitable for the proposed phonological analysis of the orthography.

It is readily admitted that this pronunciation key has several critical weaknesses. Perhaps the most limiting is the fact that this pronunciation code is a key for *reading*. The dictionary user looks up a word he does not know how to pronounce, translates the graphemes of the word into the pronunciation key provided and thereby learns how the word is said. The speller, on the other hand, is faced with a reverse process: that of transcribing spoken words into graphic form.

In addition, in the dictionary chosen the number of degrees of length by which the vowel classifications may be pronounced assumes an auditory acuity on the part of American-English speakers that certainly is not usually achieved. The seven gradations of pronunciation of the vowel classification < A >, for example, range from the long sound of this vowel phoneme as in ale to the shortened unaccented pronunciation heard in sofa. More properly, by linguistic standards, these seven degrees of length are comprised of three distinct phonemes: the long /A/ phoneme heard in ale (/A/ is often considered a diphthong by most linguists), the short /A/ phoneme heard in bat, and the unaccented vowel phoneme /ə/ as in sofa, and their allophonic variations.

These rather obvious restrictions were noted by the investigators with the realization that an accounting of them would eventually have to be made, were this pronunciation key to be used. It was decided that the advantages of the Merriam-Webster pronunciation code outweighed its disadvantages, at least for the initial stage of this research, and thus it was selected. How these limitations were resolved is treated later in the section, "The Regrouping of the Phonemic Code."

The Selection of Graphemic Options

The next step in preparing for the analysis of phoneme-grapheme correspondences in the vocabulary corpus was to select a list of graphemic options which represent each of the phonemes in the phonemic code. This selection process is considerably more complex than might be supposed. The investigators had concluded that each of the so-called "silent" letters that occur in words should be assigned to some phoneme, and this decision in itself was the basis of much discussion before the final list of graphemic options was constructed.

One of the first dilemmas encountered by the researchers was how to treat the final grapheme < E > which often functions as a diacritic marker for a preceding vowel phoneme in such words as make, life, and code. For example, < E > in the word make helps to indicate the pronunciation of the internal vowel phoneme /A/ which is represented by the grapheme < A >. Here, < E > serves as a diacritic marker, or a helping letter,[2] rather than as a silent letter. Technically, the graphemic option for /A/ in make is < A...E >.

For reasons which will be explained in a discussion of the problems involved in preparing computer programs to make phoneme-grapheme matches, an artificial phoneme category /E9/ was created for handling the grapheme < E > which in word final position "helps" to indicate the pronunciation of an internal vowel phoneme. Thus, in words containing this graphemic option, the internal vowel phoneme was assigned the graphemic option occurring in the same position, while the final occurring < E > was classified under /E9/. In make, for example, the graphemic option < A > is representative of the vowel phoneme, while the graphemic option < E > is representative of the artificial phoneme category /E9/.

Of more immediate concern, however, was the necessity to come to a decision regarding the classification of other graphemes which seemingly represent no pronounced phoneme, as exemplified by the graphemic option < B > in lamb. Here, < B > apparently serves no obvious function, since the ending consonant phoneme /M/ could be represented in lamb by only the graphemic option < M >. In this sense, < B > is an extra letter,[3] although technically the correct spelling of /M/ in lamb is the graphemic option < MB >. Considerable disagreement occurs among linguists regarding how such graphemes should be classified (Francis, 1959; Hall, 1961; Wijk, 1959). For example, is the graphemic option < U > in guard a part of the graphemic representation of the initial consonant phoneme /G/ and thus spelled < GU >, or is it more properly a part of the spelling vowel phoneme and thus spelled < UA >?

---

[2] It may be helpful to the reader to think of the grapheme < E > in this instance as a helping letter, since this is the way such graphemes are sometimes described in spelling instruction.

[3] It may be helpful to the reader to think of the grapheme < B > in this instance as an extra letter, since this is the way such graphemes are sometimes described in spelling instruction.

Even more troublesome to classify were the graphemic options representing the consonant phoneme /SH/ in words such as mission, nation, and conscience. For example, how is the graphemic option < I > classified in these words? Does this grapheme form part of a graphemic option < SSI > or < IO > in mission? Is < I > a part of the graphemic option < TI > or < IO > in nation? In conscience, is < I > a part of the graphemic option < SCI > or < IE >? Again, linguists disagree upon this point.

The decision was made to classify such graphemes in the following way: whenever single graphemic symbols occur in words which apparently do not represent a phoneme, they would be treated as a part of the graphemic representation for a phoneme. In this manner, a single consonant grapheme would be assumed to be a part of the spelling of a consonant phoneme, such as < B > being a part of the graphemic option < BT > in debt. A similar assignment was made for a single vowel grapheme of this type. For example, < U > would be considered a part of the vowel spelling < UA > in guard. Unfortunately, complete consistency with this criterion could not be maintained, and so some exceptions to this general rule will be found among the list of graphemic options used in this research.

In addition, it will be noted that some graphemic options are not listed which might be considered legitimate representations of certain phonemes. The relatively rare occurrences of some graphemic options in the orthography, however, made it wise to exclude them from the listing. It is quite true that a number of words in American-English contain rare or unique graphemic options and these words should be analyzed as well. The frequently used word of, for example, (see earlier discussion on page 12) contains unique graphemic representations of its phonemes. The investigators felt it unnecessarily cumbersome to include all the rare graphemic options that might occur in the corpus since these rare spellings have little or no phonological relationship inherent in them. Such words, as well as a number of foreign words which have been introduced into American-English, are exceptions to the alphabetic principle and should depend upon eye and hand learning.

Furthermore, many of these graphemic options were not lost to the research, since they are included in a list of "true error" words. These are words in which the computer could find no graphemic option for one or more phonemes in them. Appendix D provides a listing of the words in the corpus which were classified as "true error" words by the computer and listed separately.

The Research Design

An undertaking of this magnitude would hardly be practical were it not for the development of computer technology. Hanna and Moore required nearly three years to complete an analysis of phoneme-grapheme correspondences in 3000 words. Yet, in this research the investigators wished to analyze in even greater depth the basic phonological structure of 17,310 words in a lesser period of time. Only the analytic capacities of the modern electronic computer make such studies as this possible within a relatively short period of time.

It will be recalled that the phonological structure of spoken languages, as represented in alphabetic orthographies, is more than a matter of simple phoneme-grapheme correspondences. The investigators wished first to determine whether the position of phoneme-grapheme correspondences in syllables and the degree of stress given syllables affects how closely the American-English orthography approximates the alphabetic principle, as defined for this research. Later other relationships are examined.

It was therefore decided that each phoneme-grapheme correspondence should be classified according to three kinds of frequency and percentage tabulations:

1. A frequency and percentage tabulation of phoneme-grapheme correspondences, irrespective of phonological factors, as they occur throughout the corpus.
2. A frequency and percentage tabulation of phoneme-grapheme correspondences as they occur in initial, medial, and final positions in syllables.
3. A frequency and percentage tabulation of phoneme-grapheme correspondences in each of these positions in syllables, but including the amount of stress given to the syllables in which they occur.

In addition, it was considered necessary to produce a listing of each of the words in which these phoneme-grapheme correspondences occur, according to syllabic position and stress.

The Preparation of the Raw Data

Each factor that an investigator wishes to analyze, no matter how minute, must be written into a program--a series of instructions informing the computer which mathematical operations it must perform to produce desired data.

These instructions must be written in a specially prepared computer code which directs the computer to make the kinds of analyses the investigator wishes to have done. This code, the machine language, is normally a composite of mathematical symbols and words written in the conventional orthography, since no efficient programs have yet been developed which can be written solely in the standard orthography. In addition, in language studies where pronunciation keys are a necessary part of the information being analyzed, the symbols representing these pronunciations must be transliterated into symbols the computer is capable of handling.

The initial operation in preparing the vocabulary corpus for computer analysis was to reproduce these words on data processing cards. A special code had to be devised which transcribed the pronunciation spellings of the word list into conventional graphemic and numerical characters and certain punctuation symbols which are available on standard keypunch equipment.

A pilot study of 565 words was conducted to test the accuracy of preparing the raw data and the feasibility of the research design. Working in collaboration with the Stanford University Statistical Laboratories, under the direction of Professor Herbert Solomon, the pilot study was developed and tested at the Stanford University Computation Center. The results of this pilot study indicated that, in the main, both the coding procedure and the research design would provide the information desired when the major study was undertaken.

The task of coding the word list and placing the coded forms on data processing cards for the major study was let by bid to the Service Bureau Corporation of Palo Alto, California. Analysts were hired and trained by this firm to prepare the 17,310 word listing in coded form on sheets devised for the task. These sheets, in turn, were then given to keypunch operators who prepared data processing cards containing the coded pronunciation spelling broken into syllables showing the amount of stress given, the unsyllabicated graphemic spelling of these words, and the frequency of usage code assigned to each word. The accuracy of their work was most impressive. A later check of the error rate entailed in this phase of the study disclosed that only one-fourth of one percent of the words in the corpus were either incorrectly coded or incorrectly keypunched.

It is necessary at this point to discuss in some detail the code which was devised for the study. Computers print the output of data in the same form by which the information to be analyzed has been prepared. It may be helpful to the reader to examine the alphabetized word listing provided in Appendix A, as the listing appearing there is a copy of the computer printout.

Table 2 presents a listing of the phonemic analysis code, giving the Merriam-Webster symbol used for each pronunciation listed, the concomitant coded form used in place of this symbol, and a sample word which indicates the pronunciation represented by these symbols. The reader should attempt to understand this code if an analysis of the data to follow is to be meaningful. At first glance, the code appears somewhat

TABLE 2

PHONEMIC ANALYSIS CODE: 62 PHONEME CLASSIFICATION

Vowels				Consonants			
Coded Form	Merriam-Webster		Word Example	Coded Form	Merriam-Webster		Word Example
A	ā	as in	a̲le	B	b	as in	b̲ut
A1	a		cha̲otic	D	d		d̲ay
A2	a		ca̲re	F	f		f̲ill
A3	a		a̲dd	G	g		g̲o
A4	a̲		a̲ccount	H	h		h̲at
A5	a		a̲rm	J	j		j̲oke
A6	a		a̲sk	K	k		k̲eep
A7	a̲		sofa̲	L	l		l̲ate
E	ē		e̲ve	M	m		m̲an
E1	e		e̲vent	N	n		n̲od
E2	ē		he̲re	P	p		p̲en
E3	e		e̲nd	R	r		r̲at
E4	e̲		sile̲nt	S	s		s̲it
E5	e		make̲r	T	t		t̲o

TABLE 2--Continued

I	ī	as in	i̲ce	V	v	as in	v̲an
I3	i		i̲ll	W	w		w̲in
I4	i̲		chari̲ty	Y	y		y̲et
O	ō		o̲ld	Z	z		z̲one
O1	o		o̲bey	CH	ch		c̲h̲air
O2	o		o̲rb	HW	hw		w̲h̲at
O3	o		o̲dd	KS	ks		ethic̲s̲
O4	o̲		co̲nnect	KW	kw		g̲u̲ilt
O5	o		so̲ft	L1	'l		abl̲e̲
O6	o͞o		fo̲o̲d	M1	'm		chasm̲
O7	oo		fo̲o̲t	N1	'n		pardon̲
U	ū		cu̲be	NG	ng		sin̲g̲
U1	ū		u̲nite	SH	sh		s̲h̲e
U2	u		u̲rn	T1	th		t̲h̲in
U3	u		u̲p	T2	th		t̲h̲en
U4	u̲		circu̲s	ZH	zh		az̲ure
OI	oi		o̲i̲l	H9	--		h̲onest
OU	ou		o̲u̲t				
E9	--		late̲				

cumbersome and perhaps difficult to learn. Certain generalizations, however, may prove helpful.

It may be noted first that the phonemic analysis code is made up of the 26 letters of the alphabet and the digits 0 through 9 (excluding 8). Letters and digits are then combined to provide a coded form for each phoneme classified. It may also be noted that all letters are printed in upper-case type, this being necessary since computers print their information only in this form.

A consistency of coding was attempted for each of the types of phonemes listed. An examination of the phonemic analysis code will reveal that:

1. The "long" vowel phonemes are indicated by letters only: A, E, I, O, and U.
2. The "half-long" vowel phonemes, where designated, are represented by a grapheme and the numeral 1: A1, E1, O1, and U1.
3. The vowel phonemes classified by the circumflex diacritic mark are represented by a grapheme and the numeral 2: A2, E2, O2, and U2.
4. The "short" vowel phonemes are represented by a grapheme and the numeral 3: A3, E3, I3, O3, and U3.
5. The "italicized" short vowel phonemes are represented by a grapheme and the numeral 4: A4, E4, I4, O4, and U4.
6. Diphthongs are represented by their graphemic options: OI and OU.
7. The voiced /th/ consonant phoneme is represented by T1.
8. The unvoiced /th/ consonant phoneme is represented by T2.
9. Phonemes not included in the preceding classifications are coded by additional graphemes and numerals where they are needed.

Not included in Table 2, but a part of the code used to designate the degree of stress assigned to syllables, are the added symbols < + >, < - >, and < . >. These symbols were used in the following manner: < + > following a phonemic syllable indicates that the syllable has been given a primary stress; < - > following a phonemic syllable indicates that the syllable has been given a secondary stress; and < . > following a syllable indicates that this syllable is unaccented.

In sum, each word in the corpus was keypunched on a separate data processing card on which the following information was placed:

1. The frequency code number assigned to the word.
2. The "packed" or unsyllabicated spelling of the word in the conventional orthography.
3. The coded phonemic spelling of the word, showing its phonemic syllabication, and the amount of stress given to each syllable.
4. A several-digit serial number which identifies the word, should the data processing cards be disarranged.

It was from this information that the task of analyzing phoneme-grapheme correspondences and their approximations to the alphabetic principle in the American-English orthography was initiated.

The Preparation of the Computer Program

A contract was let by bid to Control Data Corporation, Palo Alto, California, for the preparation of a computer program by which to analyze the 17,310 word list as it had been prepared. Inasmuch as no previous computer program had been developed explicitly to provide the data desired in this study, considerable time was spent with programmers in analyzing the kind of output wanted and the most effective means of programing the raw data to produce the output.

In general, the prepared program was to instruct the computer to search the graphemic spelling of each word and to determine the frequency and percentage of occurrence in which a particular phoneme was represented by a particular graphemic option throughout the corpus.

Because certain "errors" appear later in the data, it would be useful to follow the steps a computer takes in determing which graphemic options represent the phonemes used in the lexicon. The word *problem* appears on a data processing card in this form:

1 PROBLEM P R O3 B +L E3 M 05035000

In this example, the numeral *1* appearing in the left-hand field is the frequency of usage code assigned to the word. In the right-hand field, the digits *05035000* represent the sequence number assigned to this word when it was keypunched. The portions remaining are the orthographic representation of this word and its pronunciation respelling.

The computer picks up the phoneme /P/ in the right-hand field and begins a search of the graphemic spelling of the word in the left-hand field to determine which of the graphemic options provided for the spelling of /P/ is used. < P > is the only choice possible, so the computer stores in memory one occurrence of /P/ spelled < P > in initial position in a primary accented syllable. The procedure then is followed for each of the remaining phonemes in *problem*, and a separate record is kept of each phoneme-grapheme correspondence to which all other similar correspondences in the corpus are added.

Words such as *problem*, however, are comparatively simple to analyze. There is only one possible graphemic match for each phoneme. Words such as *access* are much more difficult to analyze, as will be seen, and are sources for many of the "errors" which appear in the data.

The American-English orthography uses many doubled graphemic options to represent consonant phonemes. Note, for example, that < CC > is the graphemic option representing /K/ in *occur* when it is broken into *phonemic* syllables. In speaking this word, the syllable boundary follows the phoneme /O4/, and so the graphemic option < CC > is assigned as one of the possible spellings of the consonant phoneme /K/.

Observe, however, the difficulties encountered in assigning the appropriate graphemic options to phonemes in the word *access*. It appears on a data processing card as:

```
    4 ACCESS            A3 K +S E3 S           000093
```

The computer picks up the phoneme /A3/ in the right-hand field and matches it with the graphemic option < A > in the left-hand field. The next phoneme encountered is /K/ and at this point an ambiguity in grapheme option selection occurs. Shall it match /K/ with the graphemic option < CC > as in *occur*, or with < C > which is another legitimate graphemic option for this phoneme, as in *cat*? The problem is temporarily set aside, by noting both graphemic options and proceeding to the next phoneme /S/. Immediately another difficulty develops, since /S/ could be represented by < C > as in *cent*, but could only be assigned the graphemic option < E > if < CC > has already been assigned to /K/. Searching the listing of graphemic options for /S/, it notes that < E > is not a legitimate representation of /S/, but that < C > is. With this information, the computer then is able to return to /K/, assign to it the graphemic option < C >, then assign the second < C > to the phoneme /S/, and move on to an analysis of the next phoneme in *access*, /E3/. For this phoneme, the graphemic option < E > is the only selection possible, and thus < E > is matched with /E3/. However, a minor problem develops when the final /S/ is reached. Both the graphemic options < S > and < SS > are legitimately used for this phoneme. However, since many words end in a doubled consonant graphemic option, a rule was provided the computer which instructed it to match final occurring phonemes with any remaining graphemes, in this case assigning < SS > to /S/.

Although such analytic operations appear to be quite complicated, these operations are carried on in fractions of a second and are efficient most of the time. A trial run of the computer program, however, indicated that such instruction would not be useful for analyzing the spellings of some words in the corpus, and an effort was made to provide added instructions which would help the computer resolve ambiguities between phoneme-grapheme correspondences whenever they might occur.

To demonstrate this added procedure, and to enable the reader to understand the nature of a special group of words containing ambiguous spellings, observe the special problems encountered in matching the appropriate graphemic options to the phonemes of the word acknowledge. This word appears on a data processing card in this manner:

```
4  ACKNOWLEDGE          A4 K .N 03 L +E3 J          000139
```

The difficulty encountered in acknowledge is the selection of the appropriate spelling of the phoneme /K/. Two possible graphemic options for /K/ are < C > and < CK >. Noting this, the computer temporarily stores both options and seeks a graphemic match for the next phoneme, /N/, and thereby attempts to resolve the spelling of /K/ as was described above for the word access and its spelling of /K/ with the graphemic option of < C >. But now note a further problem: /N/ also has two legitimate graphemic options, < N > and < KN >, both of which are entirely possible in acknowledge. Observe the dilemma faced by the computer in this instance. Shall it assign the graphemic symbol < K > to /K/ so that it is spelled < CK >? Or shall it assign < K > to /N/ so that it is spelled < KN >? Such ambiguities, it turned out, are not completely resolvable.

Obviously, the computer had to be provided with instructions to make a choice whenever such ambiguities occurred. When the investigators analyzed the situation, it was noticed that three possible choices were available to the computer:

1. It could legitimately select the first possible graphemic option to be the correct representation, in this instance /K/ = < C >.
2. It could legitimately select the second possible graphemic option as the correct spelling, in this case /K/ = < CK >.
3. It could randomly select, by writing in a simple instruction, either graphemic option and assign any remaining graphemes to the next phoneme. In this instance, if it randomly selected < C > to match with /K/, it would assign < KN > as the representation of /N/.

It was apparent that a random selection would be the least productive, since there would be no way of knowing the number of times such an operation was made. Thus, either of the first two choices would provide a correct graphemic option for the phonemes involved 50% of the time. It was then noted that a simple generalization could be made, one that calls to mind the late George Zipf's "principle of least effort" (1949), and would simplify the writing of an instruction which the computer would follow. The computer was instructed to select the least number of graphemes that could legitimately represent the first phoneme it encountered in words of this type and to assign the ambiguous grapheme to the next phoneme.

The computer was also instructed to make a separate listing of the words in which this rule was applied. Thus, in acknowledge, by following this rule, < C >, would represent /K/ and < KN > would represent /N/. In effect, the first of the three possible choices was settled upon as being the most productive routine that the computer might follow.

As it happened, this particular instruction worked in over two-thirds of the cases in which it was applied. Appendix C lists the words in which these ambiguities occur. This particular problem also accounts for some error in the frequency and percentage tabulations of phoneme-grapheme correspondences. Since all "ambiguity words" were included in the data analysis, while "true error words" were not, the number of cases in which this rule did not work resulted in some rather interesting phoneme-grapheme correspondences.

An examination of the tabulated graphemic options for certain phonemes will indicate that some phoneme-grapheme correspondences occur in positions which obviously are not permissible in American-English. Thus, the graphemic option < LF > is sometimes tabulated as representing /F/ in initial position in the second syllable of compound words such as bullfrog. Such errors, however, occurred infrequently enough to have little distorting effect upon the accuracy of the frequency and percentage tabulations of graphemic representations for particular phonemes. Indeed, such words are easily spotted when the word listing is scanned and provide visual evidence of some of the orthographic peculiarities which can occur in written American-English.

Additional complications in preparing a program for the major analysis should be noted since the data reflect these problems as well. It will be recalled that a separate artifical phoneme category /E9/ was created to account for < E > which is quite commonly a part of the spelling of an internal vowel phoneme, as in same or life. At the time, programing a routine that would successfully match the graphemic option < E > to internal vowel phonemes, such as < I-E > for /I/ in life, seemed inordinately difficult and financially expensive. Both the investigators and the programers were still in the "learning" stage of applying computer technology to an intricate analysis of the American-English orthography.

Although frequency and percentage tabulations of phoneme-grapheme correspondences in the corpus are somewhat distorted by this choice, some advantages may also be observed. The many occurrences of < E > as a part of the spelling of internal vowel phonemes have been listed separately under the phoneme classification /E9/ and therefore made known the number of times this condition occurs in the 17,310 word corpus.

In addition, each word in which this condition occurs was listed under /E9/ in the separate word lists accompanying the frequency and percentage tabulations. This word list makes available a list of all the different words in which < E > was matched to /E9/ and therefore permits these words to be analyzed separately. It cannot be denied, however, that the classification of such spellings under the /E9/ category resulted in artificially high frequency and percentage tabulations for some graphemic options representing vowel phonemes.

Another limitation should also be made evident. The grapheme < H > sometimes occurs in words without an apparent phoneme to which it can be matched. Note, for example, the treatment of < H > in honor, honest, and vehement. Since this particular problem occurs in only a small number of words (41 occurrences as tabulated subsequently by the computer), it was decided not to assign < H > as a part of the graphemic representation of any consonant phoneme. As with the grapheme < E >, as a part of an internal vowel spelling but classified under /E9/, an artificial phoneme category /H9/ was established to handle this situation when it occurred. The data are not seriously affected by this classification, and it clearly simplified the writing of the computer program.

In summary, the initial computer analysis of phoneme-grapheme correspondences in the American-English orthography was prepared so that three sets of data would be available:

1. An alphabetized listing of the 17,310 words in the corpus. (See Appendix A for the complete listing of these words.)
2. A table of frequency and percentage tabulations of graphemic options for each of 62 phonemes, and classified by position in stressed and unstressed syllables.
3. A listing of the words which were analyzed, according to the occurrence of each phoneme-grapheme correspondence by position in stressed and unstressed syllables.[4]

The Regrouping of the Phonemic Code

Using a 62 phoneme classification, several limitations occur regarding what can legitimately be said of the American-English orthography's approximation to the alphabetic principle. When /E9/ is also included, it is apparent that the initial data are not as accurate as they might be. Obviously, a more realistic phoneme classification, more in accord with conventional phonemic alphabets, plus definitely assigning the grapheme < E > representing /E9/ to an internal vowel spelling, would

---

[4] Flowchart diagrams of the computer programs prepared by Control Data Corporation, Palo Alto, California, for conducting these analyses comprise Appendices E and F of the unpublished doctoral dissertation, "An Analysis of the Phonological Structure of American-English Orthography" (Stanford University, 1964) by Richard E. Hodges.

TABLE 3

PHONEMIC ANALYSIS CODE: 52 PHONEME CLASSIFICATION

Regrouped Vowel[a] Phonemes		
Coded Form	Vowel Phonemes Included	Example
A	A, A1 as in	ale
A2	A2	care
A3	A3, A4, A6	add
A5	A5	arm
E	E, E1	eve
E2	E2	here
E3	E3	end
E5	E5	maker
I	I	ice
I3	I3	ill
O	O, O1	old
O2	O2	orb
O3	O3	odd
O5	O5	soft
O6	O6	food
O7	O7	foot
U	U, U1	unite
U2	U2	urn
U3	U3	up
OI	OI	oil
OU	OU	out
SCHWA /ə/	A7, E4, I4, O4, U4	about

---

[a] Consonant phoneme code remains the same as was used in the original 62 phoneme classification. See Table 2, page 39.

give a much clearer picture of how the phonological structure is represented in the orthography.

Accordingly, an effort was made to reduce the number of vowel phoneme classifications to a more reasonable figure. Table 3 provides the classification which was made when the vowel phonemes were regrouped. It will be noted that long and short vowel phonemes are retained and that their allophonic variations are assigned to them. In addition, all weakened forms of vowel phonemes were grouped together to comprise the vowel phoneme /ə/. It will also be noticed that vowel phonemes before /R/ were retained as separate phonemes. In sum, the number of distinct vowel phoneme categories was reduced from 32 in an initial analysis to 22 in a subsequent analysis using a regrouped phonemic alphabet. This latter analysis comes much closer to giving a more concise picture of the nature of the American-English orthography and its relationships to the phonological structure of the spoken language. It is the results of this second analysis, then, upon which we will focus in Chapter 4.[5] Still retained in this second analysis is the artificial phoneme category /H9/, as well as the listing of "true error" words for which no graphemic options were provided for the particular phonemes under examination in each word.

With the development of these two complementary analyses of the American-English orthography, it was now possible to seek answers to the basic questions underlying this research: To what degree does the American-

---

[5] The reader who is interested in the results of the preliminary analysis which utilized the 32 vowel phoneme categories is referred to in Chapter IV of the unpublished doctoral dissertation of Richard E. Hodges, "An Analysis of the Phonological Structure of American-English Orthography" (Stanford University, 1964).

English orthography approximate the alphabetic principle? And, what are the relationships between the phonological structure of spoken language and its representations in the orthography?

CHAPTER 4

PHASE I: AN ANALYSIS OF THE PHONOLOGICAL STRUCTURE OF AMERICAN-ENGLISH ORTHOGRAPHY

Phase I of this research represents an attempt to make an accounting of how closely the American-English orthography approximates the alphabetic principle. The data are described and interpreted in the following way.

1. An analysis of phoneme-grapheme correspondences irrespective of other phonological factors.
2. An analysis of phoneme-grapheme correspondences in relation to their position in syllables.
3. An analysis of phoneme-grapheme correspondences in relation to their position in stressed and unstressed syllables.

Graphemic Options Representing the Phonemes of American-English

Tables 4 and 5 list the graphemic options used to represent each of the phonemes classified in this study. Certain limitations, however, should be noted regarding the legitimacy of the occurrences of some of these graphemic options in the data to follow.

Since consonant phonemes more nearly approximate the alphabetic principle than do vowel phonemes in the orthography, it would seem likely that the computer would encounter little difficulty in matching appropriate phoneme-grapheme correspondences. The reverse situation was found to be the case. The reasons for incorrect tabulations of certain consonant

TABLE 4

GRAPHEMIC OPTIONS REPRESENTING PHONEMES BY POSITION IN SYLLABLES: 22 VOWEL CLASSIFICATION

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
A				
	A as in	an gel	bass	ha lo
	A-E	ate	blade	
	AI	aid	pain	daily
	AI-E	aide	praise	
	AIGH		straight	
	AU-E		gauge	
	AY		side ways	way
	AY-E			aye
	E			cafe
	E-E		fete	matinee*
	EA		break	yea
	EI		vein	hei nous
	EI-E		seine	
	EIGH	eight	freight	weigh
	ET			beret
	EY			they
A2				
	A	ti ar a	var y	
	A-E		care	
	AI	air	hair	
	AI-E	mill ion aire		
	E		bo ler o	
	E-E	ere	there	
	EA		bear	
	EI	heir	their	
	EY	eyr ie		
A3				
	A	ad	cab	ba boon
	A-E	mad ame	bade	
	AI		plaid	
A5				
	A	arm	far	schwa
	A-E	are	large	
	AH		dah lia	bah
	AU	aunt	laugh	
	E	en core	ser geant	
	EA		heart	

TABLE 4--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
E				
	AE			al gae
	E		sen ior	
	E-E	eve	theme	
	EA	eat	beat	sea
	EA-E	ease	please	
	EE	eel	feel	bee
	EE-E		geese	
	EI		ceil ing	nei ther
	EI-E		seize	
	EO			peo ple
	EY			key
	I	ser ies	mo tif	ski
	I-E	na ive	e lite	
	IE		chief	
	IE-E		piece	
	OE			a moe ba
E2				
	E	heer i al	her o	
	E		here	
	EA	ear	near	
	EE	auc tion eer	deer	
	EI		weird	
	I		sou ve nir	
	IE	cash ier	pier	
	IE-E		fierce	
E3				
	A	an y way	man y	
	A-E	or ange		
	AE	aes thetic		
	AI		said	
	AY		says	
	E	end	bet	be ret
	E-E	edge	ledge	
	EA		head	
	EA-E		cleanse	
	EI		heif er	
	EO		leo pard	
	IE		friend	
	U		bur y	

TABLE 4--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
E5				
	A	cow ard	al tar	
	E	bak er	af ter	sou ve nir
	E-E	trav erse		
	I	aff ir ma tion	na dir	
	O	hon or	hu mor	
	OU	glam our		
	U	treas ur y	sur vey	
	U-E	meas ure	in jure	
	Y	sat yr	mar tyr	
I				
	AI-E	aisle		
	AY			kay ak
	EI		stein	ka lei do scope
	EIGH		height	
	EY			geyser
	EY-E			eye
	I	ic ing	find	chi na
	I-E	ice	like	
	EI		tried	pie
	IGH		night	high
	OY			coy ote
	UY			buy
	Y		typ ist	by
	Y-E		type	dye*
I3				
	A	spin ach		
	A-E	sen ate	lan guage	
	AI	murr ain	cap tain	
	AY			yes ter day
	E		prett y	a do be
	E-E		priv i lege	
	EA			guin ea
	EA-E	mil eage		
	EE		been	coff ee
	EI	for eign	for feit	
	EY			don key
	I	in	did	ap ri cot
	I-E	fa vor ite	give	
	IA-E	marr iage		
	IE	var ied	mis chief	mo vie
	IE-E		sieve	
	O		wom en	
	U		bus y	

TABLE 4--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
I3--Continued				
	U-E	min ute		
	UI		build	gui tar
	Y	lar ynx	myth	ba by
	Y-E		a poc a lypse	
O				
	AU			hau teur
	AU-E		mauve	
	EAU			beau
	EW			sew
	O	on ly	both	so
	O-E	ode	code	
	OA	oak	coat	co coa
	OA-E		coarse	
	OE			doe
	OH	ohm		
	OO		door	
	OU		four	
	OU-E		course	
	OUGH			dough
	OW	own	sown	bow
	OW-E			owe
OI				
	OI	oil	coil	poi son
	OI-E		voice	
	OY			boy
	OY-E		gar goyle	
OU				
	OU	out	bout	thou sand
	OU-E	ounce	house	
	OUGH		drought	bough
	OW	owl	howl	tow el
	OW-E		browse	
O2				
	A	al so	hall	wa ter
	A-E		false	
	AO		ex traor di nar y	
	AU	auc tion	laud	fau cet
	AU-E		cause	
	AUGH	aught	caught	daugh ter
	AW	awn ing	dawn	law
	AW-E			awe

TABLE 4--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
02--Continued				
	O	or	cord	
	O-E		horse	
	OA		broad	
	OUGH	ought	bought	
03				
	A		wad	
	O	odd	dot	
	O-E		dodge	
	OW		knowl edge	
05				
	O	off	log	
	O-E		gone	
06				
	EU		sleuth	ma neu ver
	EW		shrewd	threw
	O		whom	who
	O-E		lose	
	OE			can oe
	OO		boot	too
	OO-E	ooze	choose	
	OU		soup	you
	OU-E		route	
	OUGH			through
	U		truth	tu na
	U-E		rude	
	UE			blue
	UI		fruit	
	UI-E		juice	
	UO			buo yant
07				
	EU		pleur i sy	
	O		wom an	to day
	OO		wood	
	OU		could	tam bou rine
	U		put	plu ral i ty
	U-E		sure	
ə				
	A	pal at a ble	can vas	ca nal
	A-E	me di ate	pi rate	
	AI	vill ain	cer tain ty	moun tai nous

TABLE 4--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
ə --Continued				
	AU			ep au let
	E	reb el	an gel	
	E-E	sci ence	li cense	
	EA	pag eant	ser geant	
	EI	sov er eign ty		for fei ture
	EO	pig eon	dun geon	
	EOU		cou ra geous	
	I	civ il	pen cil	mu ti ny
	I-E	frag ile	en gine	
	IA			par lia ment
	IE	pro fic ient	an cient	mis chie vous
	IE-E	om nisc ience	pa tience	
	O	id i om	car ton	co coon
	O-E	liss ome	wel come	
	OI-E		tor toise	
	OU	pi ous	e nor mous	cou ra geous
	U	me di um	sub mit	u ku le le
	U-E	ferr ule		co que try
	Y	eth yl	i dyl	por phyry
U				
	EAU			beau ty
	EU		feud	feu dal
	EU-E		deuce	
	EW		news	new
	EW-E			ewe
	IEU			lieu
	IEW			view
	U	un ion	re fus al	fu ry
	U-E	use	cute	
	UE	val ued		cue
	UI		suit	mui sance
U2				
	E	err	her	
	E-E	co erce	verse	
	EA	earn	learn	
	EA-E		hearse	
	EU		am a teur	
	I	irk	bird	
	I-E		dirge	
	O		word	
	O-E		worse	
	OU		jour ney	
	OU-E		scourge	

TABLE 4--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
U2--Continued				
	U	urn	burn	
	U-E	urge	curse	
	Y		myr tle	
U3				
	O	ov en	son	
	O-E		come	
	OO		blood	
	OU		touch	
	U	up	cup	
	U-E		budge	

[*] Erroneous phoneme-grapheme match due to computer error.

TABLE 5

GRAPHEMIC OPTIONS REPRESENTING PHONEMES BY POSITION IN SYLLABLES: 30 CONSONANT CLASSIFICATION

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
B				
	B as in	bad		
	BB			ebb
CH				
	C	cell o		
	CH	chart	arched	arch
	T	cul ture		rit u al
	TCH			watch
	TI	ques tion		
D				
	D	dad	ads	ad
	DD		odds	add
	LD			could
F				
	F	farm	aft	beef
	FF			cuff
	FT			oft en
	GH			rough
	LF			calf
	PH	phone	sphere	graph
G				
	G	gap	hogs head	beg
	GG			egg
	GH	ghost		burgh
	GUE			rogue
	X			ex act
H				
	H	hot		
	WH	who		
HW	WH	what		
H9				
	H	heir		

TABLE 5--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
J				
	D as in	ar du ous		ed u cate
	DG			budg et
	DI	cor dial		
	DJ	a djust		adj ec tive
	G	gem		leg end
	GG			ex agg er ate
	GI	re gion		
	J	jam		proj ect
K				
	C	cab	act	arc
	CC			occ u py
	CCH			sacch a rin
	CH	chorus	school	ech o
	CK		cracked	back
	CQ			acq ui esce
	K	keep	ski	leak
	KH	khaki		
	LK			talk
	Q			liq uid
	QU	bou quet		
	SC	vi scount		
	X			ex it
KS				
	CS			eth ics
	X		fixed	
KW				
	CQU	a cquaint		
	QU	quilt	squat	
L				
	L	lad	cliff	ail
	LL		o ver alls	all
	LN			kiln
	SL	i sland		
Ll				
	AL			med al
	EL			ea sel
	IL			e vil
	LE		mea sles	a ble
	OL			pis tol

TABLE 5--*Continued*

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
M				
	GN as in			phlegm
	LM	a lmond	alms	calm
	M	man	imp	am
	MB			lamb
	MM			mumm y
	MN			hymn
Ml				
	M			Chasm
N				
	GN	gnat		sign
	KN	knee		
	MN	mne mon ics		
	N	nap	and	man
	NN			inn
	PN	pneu mo ni a		
NG				
	N		bank	an chor
	NG		length	long
	NGUE			tongue
Nl				
	EN			shor ten
	IN			ba sin
	ON			par don
P				
	P	pad	apt	cap
	PP		stopped	app le
R				
	R	rap	arm	far
	RH	rhyme		
	RR			purr
	WR	wrap		
S				
	C	cent	paced	ac id
	PS	psalm		
	S	sit	ask	bus
	SC	scene		musc le
	SCH	schism		

TABLE 5--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
S--Continued				
	SS as in		de pressed	boss
	ST			list en
	SW	sword		
	Z	ec ze ma		waltz
SH				
	C	o ce an ic		de fic ient
	CE	o cean		
	CH	chef		ech e lon
	CI	so cial		speci al
	S	sugar		
	SC	cre scen do		fasc ism
	SCH	schwa		
	SCI	con scious		lusci ous
	SH	ship	fin ished	wish
	SI	pen sion		
	SS			press ure
	SSI			missi on
	T	ne go ti ate		i nit i ate
	TI	ac tion		i niti al
T				
	BT			debt
	CHT			yacht
	CT			in dict
	ED			paced
	PT	pto maine		receipt
	T	tab	its	bet
	TH	thyme		
	TT			putt
	TW	two		
T1				
	TH	thaw		bath
T2				
	TH	the	clothes	moth er
V				
	F			of
	LV		calves	
	V	vat	wives	civ ic
W				
	OU as in	biv ouac		
	U	liq uid	sweet	
	W	we		

TABLE 5--Continued

Phoneme	Graphemic Option	Word Examples		
		Initial	Medial	Final
Y				
	E	a zal ea		
	I	sen ior	fiord	
	Y	yes		
Z				
	CZ	czar		
	ES			goes
	S	de sign	chasm	as
	SC	di scern		
	SS	de ssert		huss y
	Z	zoo		adz
	ZZ			buzz
ZH				
	G	re gime		
	S	u sual		meas ure
	SI	fu sion		vision

phoneme-grapheme correspondences involves discussing in some detail certain peculiar characteristics of the American-English orthography.

A number of letters of the alphabet are extraneous, in the sense that other graphemes more justifiably represent the phonemes of the spoken language. The grapheme < X > is one such extraneous graphic symbol, but with an additional complicatinn. In American-English orthography, < X > is often used to represent two phonemes at one time in some words. Consider the function of < X > in *exact*. In this word, < X > represents the phoneme /G/ which ends the first syllable and likewise represents the phoneme /Z/ which initiates the second syllable of *exact*. Although the number of words in which < X > performs a dual function is not excessive in relation to the total lexicon, it does occur frequently enough to have a bearing on the tabulations of graphemic representations for both /G/ and /Z/.

The computer can use the grapheme < X > only once in assigning it to a particular phoneme. For example, once < X > has been assigned to /G/ in *exact*, the computer then is not able to assign it to /Z/. < X > is tabulated as being a graphemic option for /G/ in the final position of an unaccented syllable in *exact*. Similar problems are encountered for /K/ and /S/ in *extra*, in which < X > represents both phonemes. In all such instances, then < X > is assigned to a single consonant phoneme when, in actuality, this grapheme is representing two phonemes at one time.

An even more troublesome phoneme-grapheme relationship is found in attempting to assign appropriate graphemic options to the consonant phoneme /SH/, as was previously noted on page 34 in Chapter 3. A review of the literature which treats this phonemic category provides no ready answer to how /SH/ is usually represented in the orthography (Francis, 1958; Greene, 1954; and Wijk, 1959). Consider the graphemic options for /SH/ in various words. In *social* and *special*, does < C > or < CI > represent /SH/? In *mission*, is /SH/ represented by < SS > or < SSI >? And, in *action* and *initial*, is the correct graphemic option < T > or < TI > for the phoneme /SH/?

It will be recalled that the investigators had decided to handle any single graphemic symbols which apparently do not represent a phoneme as though they were a part of a graphemic option for a phoneme. Thus, < U > in *guard* would be assigned as a part of the vowel spelling < UA >, and < B > in *debt* would be treated as a part of the consonant spelling < BT >. In treating /SH/, however, an exception to this rule was made. Graphemic options were selected which are combinations of both consonant and vowel graphemes, as in < CE >, < CI >, < SCI >, < SI >, < SSI >, and < TI >, while some graphemic options were selected which include only a consonant grapheme, as in < C >, < CH >, < S >, < SCH >, < SH >, < SS >, and < T >. This kind of eclectic decision-making was felt to be necessary lest the large numbers of words in which this phoneme occurs be set aside as "true error" words and the information regarding the other phoneme-grapheme correspondences in these words lost to the analysis.

Even so, the computer on occasion could not make appropriate phoneme-grapheme matches in words of this type, and some erroneous correspondences occur in the data. It becomes apparent that /SH/ is one of the few consonant phonemes which is difficult to spell, especially when it occurs in unaccented syllables.

It may have been noted that a part of the problem of classifying < X > as a graphemic option for consonant phonemes involves phonemes which are separated by syllable boundaries. This situation becomes even more apparent in the classification of doubled graphemic options for certain consonant phonemes, as in /K/ in acclaim, /L/ in bullfrog, and /SH/ in assure.

A pronunciation of acclaim produces the utterance of /K/ as initiating the second syllable of this word. Note, however, that in this case the grapheme < C > occurs in succession in acclaim. An ambiguity develops which the computer handles in the same manner as other ambiguous spellings. Thus, finding no graphemic option < AC > for the phoneme of the first syllable, it assigns < CC > to /K/, which is a legitimate option for this phoneme, but not in initial position in syllables.

Likewise, in compound words such as bullfrog, where the first syllable ends in a doubled consonant graphemic option, the computer cannot make the appropriate assignment of < LL > to /L/ at the end of the first syllable. Nor can it assign < F > to /F/, since < LF > is also an appropriate graphemic option for this phoneme, as in half. Thus, following the rule established to resolve ambiguity among phoneme-grapheme correspondences, it assigns < LF > to /F/, resulting in a spelling of this phoneme in a position of a syllable where it does not legitimately occur in the orthography.

It will be recalled from page 7, Chapter 1, that vowel phonemes form the nucleus of syllables, and that consonant phonemes are then added to them in either or both initial and final positions. As a matter of fact, the consonant-vowel-consonant syllable pattern is predominant in American-English (Miller, 1951, p. 791). In this paradigm, vowel phonemes are considered always to occupy medial position in syllables, while consonant phonemes would occur only in initial or final positions. This syllable construction is usually referred to as CVC patterning.

In analyzing phoneme-grapheme relationships in syllables, the investigators had hoped to use the CVC pattern as the criterion upon which to determine the effects of syllabic position of phoneme-grapheme correspondences, Computer programming, however, can be, and often is, a complex undertaking which requires numerous subroutines of instructions to be written when especially involved information is desired. The programing of phoneme-grapheme relationships in a CVC patterned syllable turned out to be this kind of programing operation.

Consider the position of consonant and vowel phoneme-grapheme correspondences in the words string and strict. Under the CVC paradigm, these correspondences would appear as < STR I NG > and < STR I CT >, with all the consonant correspondences occurring in initial and final positions, and the vowel correspondence in medial position in these monosyllabic words. Thus, there would appear a cluster of consonants in these positions.

It would seem to be a simple matter to instruct the computer to assign all consonant clusters to either an initial or final position in syllables. However, each additional consonant phoneme occurring after the first one in such clusters involves a "branching" routine, or the development of a series of subinstructions to instruct the computer to make these assignments. For each additional consonant phoneme in clusters, the writing of instructions multiplies in a geometric progression. "Branching" routines, therefore, are especially expensive to develop.

Treating each phoneme in syllables as a separately occurring item, however, is less complex to program and therefore less expensive. As a consequence, the investigators decided partially to abandon the CVC pattern as a criterion for analyzing phoneme-grapheme correspondences by syllabic position. In turn, only the initially occurring phoneme would be considered to be in initial position in syllables, and only the final occurring phoneme would be assigned to final position in syllables. All other phonemes between these two would be considered to be in medial position in syllables. Using string as an example, /S/ would be represented by < S > in initial position, /NG/ would be represented by < NG > in final position, and the other phoneme-grapheme correspondences, < T >, < R >, and < I >, would be considered as having occurred medially.

In interpreting the data to follow, it will be helpful to remember that tabulations of consonant phoneme-grapheme correspondences in medial position in syllables have resulted from having made this decision. As a matter of fact, such information is not necessarily misleading. The treating of several consonants in sequence as clusters is primarily a task the pupil learns as a part of the act of reading. From the point of view of spelling instruction, it is actually more beneficial to make appropriate graphemic matches for each phoneme individually in such words. Only when the word becomes a part of the pupil's reading vocabulary, after he has mastered making the appropriate phoneme-grapheme matches, does an analysis of such words by consonant clusters have relevance.

Another controversial point concerns assigning the graphemic representation < ES > to the consonant phoneme /Z/. This particular option was assigned to /Z/ in order to handle effectively the phoneme-grapheme correspondences in such words as *does* and *goes*. By virtue of the attempt to assign all vowel graphemes to vowel phonemes, and similarly so for consonant phoneme-grapheme correspondences, /Z/ would be represented by < S > in each of these words, with the grapheme < E > being attributed to a part of the vowel phoneme's graphemic representation. In order to avoid the possibility of ambiguity errors in the large number of words in which this particular use of /Z/ occurs, it was considered more effective to assign < ES > as an option for this phoneme.

A further limitation should be noted which has specific inference to an interpretation of vowel phoneme-grapheme correspondences. The partial abandonment of CVC patterning of syllables creates the following possible syllabic patterns:

Initial------medial------medial------medial------final, as in S T R I NG
Initial------------medial------medial------------final, as in S T I NG
Initial------------------medial------------------final, as in S I T
Initial--final, as in D O
Initial   or   final, as in A

Thus, a singly occurring vowel phoneme which also is a syllable may be considered to be in either initial or final position. Logically, such a vowel phoneme would be tabulated as an initially occurring phoneme, since it is the first and only one which the computer analyzes in a syllabic environment. However, such vowel phonemes are usually pronounced in a long form (e.g., /O/ in obey) and the so-called "open" syllable is one in which the ending phoneme is a vowel. Because the investigators felt it wise to retain "open" syllable features of singly occurring vowel phonemes, it was decided to tabulate these phonemes as though they occur in a final position in syllables, even though strict adherence to CVC syllable patterning would indicate that they should be tabulated as occurring in medial position in syllables.

The Regrouping of Vowel Phonemes

As was discussed previously in Chapter 3, an attempt was made to synthesize the 62 phoneme code into a code more closely related to conventional phonemic alphabets. This procedure involved regrouping the vowel phoneme classifications as listed in the pronunciation key to the Merriam-Webster's New Collegiate Dictionary, sixth edition. Table 3, found on page 50ff in Chapter 3, presents the phonemic alphabet which resulted from the regrouping procedure.

In addition to reducing the number of vowel phonemes from 32 to 22, the artificially created phoneme category /E9/ was eliminated, and each occurrence of the < E > was assigned as part of the graphemic representation of some vowel phoneme. This procedure enabled the investigators to obtain a more accurate interpretation of the frequency and percentage tabulations of the graphemic options representing the vowel phonemes. The data clearly point out how this refinement both increased and decreased the degree to which the vowel phonemes approximate the alphabetic principle as they occur in their respective positions in syllables.

It should be noted, however, that even this way of treating < E >

is not entirely accurate. A certain amount of arbitrary interpretation is necessary to stipulate that < E > is always used in this instance as a part of a graphemic option for an internal vowel phoneme. Some argument can be made that < E > is equally a part of the graphemic option for the immediately preceding consonant phoneme /S/ in choice. Might it not indicate that the grapheme < C > represents /S/ in this and other similar environments since the vowel grapheme < OI > can represent only the phoneme /OI/?

A similar claim can be made for final < E > in words which end with the consonant phoneme /V/, as in have or give. Notice that in these words the grapheme < E > does not signal a lengthening of the internal vowel phoneme. And consider, as well, that American-English does not use the grapheme < V > to end words in the orthography. Despite such considerations, the resulting regrouped data present a much clearer picture of how closely the orthography approximates the alphabetic principle as defined in this research. The reader should infer that those graphemic representations of vowel phonemes which include < E > as a part of the option contain a consonant phoneme between the two parts of the option. These are presented in the tables to follow as < A-E >, and other similar configurations, in which the hyphen signals any consonant phoneme-grapheme correspondence occurring in that environment.

In the process of preparing the program for the regrouped phoneme analysis, it was felt advisable to refine any other factors which seemed to be erroneous in the initial analysis. One such instance was the classification of < L > as one of the graphemic options for the syllabic consonant /Ll/. In examining the word listings which accompanied the tables of frequency and percentage tabulations of phoneme-grapheme correspondances, it was noted that each occurrence of < L > for this phoneme was in actuality < LE >, as in able or circle. Obviously, < E > was not a part of the spelling of an internal vowel phoneme in such words; rather, it was quite logical to assume that < E > was a part of the spelling of /Ll/, a syllabic consonant. Consequently, the graphemic option < L > was changed to < LE > for this phoneme when the regrouped phoneme analysis was made and, in the process, some additional words were picked up from the "true error" list and included as part of the analysis.

A brief statement concerning the "true error" list is appropriate here. Appendix D lists those words out of the original 17,310 that were classified as errors by the researchers because of key punching mistakes or listed by the computer as true error because of several causes as stated. These errors amounted to 301 and thus the net list of words in the corpus analyzed was 17,310 minus 301, or 17,009 words.

Frequency and Percentage Tabulations of Phoneme-Grapheme Correspondences

As has been discussed, the "ideal" alphabetic orthography would always attain the alphabetic principle. This orthography would enable the user to transcribe accurately the phonemes of the spoken language into the graphemes of the written language simply on the basis of making a direct one-to-one correspondence between them, without having to take into account other factors. The ideal alphabetic orthography would not be influenced by such matters as the position of the phoneme-grapheme correspondence in syllables and words, other than requiring the writer to write the graphemes in the same sequence as the phonemes which comprise the spoken word. Nor would the user of the ideal alphabetic orthography need to be concerned necessarily with the influences of syllabic stress on phoneme-grapheme relationships.

TABLE 6

ADDITIONAL OCCURRENCES OF CERTAIN CONSONANT PHONEMES WHEN < L > IS CHANGED TO < LE > AS A GRAPHEMIC OPTION FOR /L1/

Consonant Phonemes Affected	Original Frequency Tabulation	Revised Frequency Tabulation	Increase in Frequency
B	2,303	2,306	3
D	3,691	3,703	12
G	1,338	1,342	4
K	4,712	4,714	2
L1	651	662	11
M	3,501	3,503	2
N	7,656	7,662	6
NG	615	616	1
P	3,449	3,455	6
R	9,390	9,394	4
S	6,326	6,328	2

The American-English orthography, as has been pointed out repeatedly, does not obtain so high a correlation between the spoken and written language that the alphabetic principle ideally is attained for all phoneme-grapheme correspondences. However, the orthography approximates this principle, at least to the extent that some phonemes have only one graphemic representation while some graphemic options predominantly represent certain phonemes in certain positions in stressed and unstressed syllables. Given this delimitation, the mean percentage of all phoneme-grapheme correspondences indicates that the orthography is alphabetic at least four-fifths of the time, or an average of approximately 80% (Moore, 1951).

The odds that a writer will choose an appropriate graphemic option for a particular phoneme 80% of the time is, of course, quite probabilistic since some phoneme-grapheme correspondences are much less consistent than this percentage figure suggests. Particularly among the vowel phonemes are the problems of correct graphemic selection most evident. However, many writers of American-English do spell words with considerable facility, apparently because they have intuitively sorted out other factors which help them to make correct phoneme-grapheme correspondences. It was a part of the major purpose of this research to determine the degree to which certain factors of the phonological structure underlying the orthography are possible aids to making correct phoneme-grapheme matches when a knowledge of the alphabetic principle as defined is applied.

Even with this kind of delimitation, that is, restricting the alphabetic principle to occurrences of phoneme-grapheme correspondences in given positions in stressed and unstressed syllables, it is obvious that this principle cannot completely be attained. Some standard or criterion needs to be established to which each phoneme and its graphemic options can be compared in terms of how closely these correspondences approximate the alphabetic principle. Since the Moore research has indicated that such an approximation is roughly 80%, this percentage figure will be used as the criterion to determine the degree of consistent correspondence between phonemes and their graphemic representations. The 80% criterion can, in short, serve as a useful measure of how closely each phoneme approximates the alphabetic principle in terms of such phonological characteristics as phoneme-grapheme correspondences tabulated by their occurrences in given positions in stressed and unstressed syllables.

The first factor to be considered, then, is the degree to which phoneme-grapheme correspondences approximate the alphabetic principle, in terms of the 80% criterion, irrespective of positional and stress factors. Tables 7 and 8 summarize these data for the 22 vowel classification and the 30 consonant classification. In each table, the following information is provided:

1. The phoneme classification.
2. The graphemic options representing this phoneme.
3. The frequency of occurrence of each phoneme-grapheme correspondence in the corpus.
4. The percentage of occurrence that each phoneme-grapheme correspondence attains in comparison to all the other graphemic options for a given phoneme.

22 vowel classification. An examination of the data in Table 7 reveals that only six of the 22 vowel phonemes, or 27%, /A3/, /A5/, /E3/, /03/, /05/, and /U3/, equal or exceed the 80% criterion. As would be expected, these phonemes are primarily the short vowel phonemes and their occurrences before /R/.

The 16 vowel phonemes falling below the 80% criterion are the long vowel phonemes /A/, /E/, /I/, /0/, and /U/; the long vowel phonemes before /R/ as in /A2/, /E2/, /02/, and /U2/; the short vowel phoneme /I3/; /E5/ (which is the short vowel /E5/ before /R/); the dipthongs /OI/ and /OU/; the vowel phonemes /06/ and /07/; and the vowel phonemes occurring in unaccented syllables as symbolized by /ə/. The vowel phoneme having the highest percentage of regularity in spelling is /05/ spelled < O > 96.85% of the time, and the lowest vowel phoneme-grapheme correspondence is /ə/ spelled < E > 26.79% of the time. Of the 41,135 regrouped vowel phonemes tabulated, out of a total of 108,585 phonemes in 17,009 words, 62.27% of

TABLE 7

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH: 22 VOWEL CLASSIFICATION

Phoneme	Grapheme	f	%
/A/		2248	
	A	1002	44.57
	A-E	790	35.14
	AI	208	9.25
	AY	131	5.82
	AI-E	18	.80
	EIGH	18	.80
	E	16	.71
	EA	14	.62
	EI	14	.62
	EY	14	.62
	ET	9	.40
	E-E	6	.26
	AIGH	4	.17
	EI-E	2	.08
	AU-E	1	.04
	AY-E	1	.04
/A2/		220	
	A	64	29.09
	A-E	50	22.72
	AI	46	20.90
	E-E	34	15.45
	EA	13	5.90
	EI	5	2.27
	E	4	1.81
	AI-E	3	1.36
	EY	1	.45
/A3/		4340	
	A	4192	96.58
	A-E	147	3.38
	AI	1	.02
/A5/		580	
	A	518	89.31
	A-E	31	5.34
	EA	18	3.10
	E	5	.86
	AH	4	.68
	AU	4	.68
/E/		2538	
	E	1765	69.54
	EE	249	9.81
	EA	245	9.65
	E-E	62	2.44
	I-E	44	1.73
	I	38	1.49
	IE	33	1.30
	EA-E	30	1.18
	IE-E	23	.90
	EI	16	.63
	EE-E	9	.35
	EI-E	6	.23
	EY	6	.23
	AE	5	.19
	OE	5	.19
	EO	2	.07
/E2/		198	
	E	64	32.32
	EA	49	24.74
	EE	36	18.18
	E-E	27	13.63
	IE	14	7.07
	I	3	1.51
	IE-E	3	1.51
	EI	2	1.01
/E3/		3646	
	E	3316	90.94
	EA	139	3.81
	A	94	2.57
	E-E	79	2.16
	AI	4	.10
	IE	4	.10
	EO	3	.08
	U	2	.05
	A-E	1	.02
	AE	1	.02
	AY	1	.02
	EA-E	1	.02
/E3/--Continued			
	EI	1	.02
/E5/		2170	
	E	1666	76.77
	O	268	12.35
	A	168	7.74
	U	31	1.42
	U-E	23	1.05
	I	8	.36
	Y	4	.18
	E-E	1	.04
	OU	1	.04
/I/		1482	
	I-E	555	37.44
	I	554	37.38
	Y	211	14.23
	IGH	88	5.93
	IE	26	1.75
	Y-E	23	1.55
	EY-E	7	.47
	EI	6	.40
	AY	3	.20
	EIGH	3	.20
	UY	3	.20
	AI-E	1	.06
	EY	1	.06
	OY	1	.06
/I3/		7815	
	I	5346	68.40
	Y	1801	23.04
	I-E	339	4.33
	A-E	187	2.39
	EY	40	.51
	IE	29	.37
	UI	16	.20
	AI	15	.19
	EI	11	.14
	EE	6	.07

TABLE 7--Continued

Pho-neme	Graph-eme	f	%
/I3/--Continued			
	E	5	.06
	A	4	.05
	IA-E	3	.03
	U	3	.03
	U-E	3	.03
	AY	1	.01
	E-E	1	.01
	EA	1	.01
	EA-E	1	.01
	IE-E	1	.01
	O	1	.01
	Y-E	1	.01
/O/		2587	
	O	1876	72.51
	O-E	370	14.30
	OA	126	4.87
	OW	124	4.79
	OU	29	1.12
	OE	13	.50
	OU-E	10	.38
	OO	9	.34
	OUGH	8	.30
	EAU	6	.23
	OH	4	.15
	AU	3	.11
	EW	3	.11
	OA-E	3	.11
	AU-E	2	.07
	OW-E	1	.03
/OI/		149	
	OI	92	61.74
	OY	48	32.21
	OI-E	8	5.36
	OY-E	1	.67
/OU/		406	
	OU	227	55.91
	OW	119	29.31
	OU-E	54	13.30
	OUGH	4	.98
	OW-E	2	.49
/O2/		767	
	O	312	40.67
	A	165	21.51
	AU	146	19.03
	AW	75	9.77
	O-E	17	2.21
	OUGH	15	1.95
	AUGH	12	1.56
	OA	9	1.17
	AU-E	9	1.17
	A-E	3	.39
	AO	2	.26
	AW-E	2	.26
/O3/		166	
	O	1558	93.74
	A	80	4.81
	O-E	20	1.20
	OW	4	.24
/O5/		127	
	O	123	96.85
	O-E	4	3.14
/O6/		453	
	OO	173	38.18
	U	93	20.52
	O	37	8.16
	U-E	34	7.50
	OU	29	6.40
	EW	22	4.85
	UE	16	3.53
	OO-E	12	2.64
	O-E	12	2.64
	UI	6	1.32
	EU	4	.88
	OE	4	.88
	UI-E	4	.88
	OU-E	3	,66
	OUGH	2	.44
	UO	2	.44
/O7/		368	
/O7/--Continued			
	U	200	54.34
	OO	114	30.97
	OU	25	6.79
	O	17	4.61
	U-E	11	2.98
	EU	1	.27
/ə/		6013	
	O	1611	26.79
	A	1438	23.91
	I	1347	22.40
	E	763	12.68
	OU	336	5.58
	U	297	4.93
	E-E	101	1.67
	Y	23	.38
	IE	22	.36
	O-E	21	.34
	EO	10	.16
	AI	9	.14
	EOU	8	.13
	I-E	6	.09
	IE-E	4	.06
	U-E	4	.06
	A-E	3	.04
	EA	3	.04
	EI	2	.03
	IA	2	.03
	OI-E	2	.03
	AU	1	.01
/U/		1188	
	U	814	68.51
	U-E	256	21.54
	EW	38	3.19
	EU	28	2.35
	UE	27	2.27
	UI	8	.67
	IEW	6	.50
	EAU	5	.42
	IEU	4	.33
	EU-E	1	.08
	EW-E	1	.08

TABLE 7--Continued

Pho-neme	Graph-eme	f	%	Pho-neme	Graph-eme	f	%	Pho-neme	Graph-eme	f	%
/U2/		787		/U2/--Continued				/U3/		1410	
	E	313	39.77		U-E	10	1.27		U	1212	85.95
	U	203	25.79		EU	6	.76		O	112	7.94
	I	104	13.21		EA-E	2	.25		OU	30	2.12
	O	53	6.73		Y	2	.25		O-E	26	1.84
	E-E	41	5.20		I-E	1	.12		U-E	23	1.63
	EA	29	3.68		O-E	1	.12		OO	7	.49
	OU	21	2.66		OU-E	1	.12				

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

the time they are represented in the orthography by a particular graphemic option for each of the 22 phonemes.

30 consonant classification. Consonant phonemes are more consistently represented by a particular graphemic option than are the vowel phonemes, as is indicated in Table 8.

Nineteen, or 63%, of the 30 consonant phonemes classified equal or exceed the 80% criterion. The mean percentage of consistent spellings of the 30 consonant phonemes is 83.99%. It is observed that the consonant phoneme /HW/, the syllabic consonant phoneme /M1/, and the voiced and unvoiced consonant phonemes /T1/ and /T2/ have only one graphemic option and are therefore wholly consistent with the alphabetic principle, while the consonant phoneme /ZH/ is represented in the orthography by its most frequently used graphemic option, < SI >, only 49.01% of the time.

Further analysis discloses that those consonant phonemes falling below the 80% criterion usually share a highly used graphemic option among them. Note that /K/ is spelled < C > 73.25% of the time, while /S/

TABLE 8

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH: 30 CONSONANT CLASSIFICATIONS

Pho-neme	Graph-eme	f	%
/B/		2303	
	B	2239	97.22
	BB	63	2.73
	PB	1	.05
/CH/		564	
	CH	313	55.49
	T	175	31.03
	TCH	61	10.81
	TI	13	2.33
	C	2	.35
/D/		3691	
	D	3611	97.83
	DD	74	2.00
	LD	6	.17
/F/		2019	
	F	1580	78.25
	PH	242	12.02
	FF	177	8.76
	LF	9	.44
	GH	8	.39
	FT	3	.14
/G/		1338	
	G	1178	88.04
	GG	67	5.00
	X	42	3.17
	GUE	21	1.56
	GU	19	1.42
	GH	10	.74
	TG	1	.07
/H/		778	
	H	762	97.94
	WH	16	2.06
/HW/		89	
	WH	89	100.00
/H9/		41	
	H	41	100.00
/J/		982	
	G	647	65.88
	J	218	22.24
	DG	51	5.19
	D	32	3.25
	DJ	16	1.62
	GI	14	1.42
	GG	2	.20
	DI	2	.20
/K/		4712	
	C	3452	73.25
	K	601	12.75
	CK	290	6.15
	CH	142	3.01
	X	80	1.75
	CC	76	1.61
	QU	27	.57
	Q	20	.42
	LK	14	.29
	CQ	3	.06
	KH	3	.06
	SC	3	.06
	CCH	1	.02
/KS/		271	
	X	245	90.41
	CS	26	9.59
/KW/		196	
	QU	191	97.45
/KW/--Continued			
	CQU	5	2.55
/L/		5389	
	L	4894	90.81
	LL	489	9.07
	SL	5	.11
	LN	1	.01
/Ll/		651	
	LE	620	95.23
	EL	19	2.91
	IL	7	1.07
	AL	4	.61
	OL	1	.18
/M/		3501	
	M	3302	94.31
	MM	140	3.99
	MB	27	.77
	LM	17	.48
	GM	8	.22
	MN	7	.23
/Ml/		97	
	M	97	100.00
/N/		7656	
	N	7452	97.33
	NN	127	1.65
	KN	41	.53
	GN	32	.41
	PN	3	.07
	MN	1	.01
/NG/		615	
	NG	362	58.86

TABLE 8--Continued

Pho-neme	Graph-eme	f	%
/NG/--Continued			
	N	251	40.81
	NGUE	2	.33
/N1/		128	
	EN	84	65.62
	ON	41	32.04
	IN	3	2.34
/P/		3449	
	P	3296	95.56
	PP	153	4.44
/R/		9390	
	R	9119	97.11
	RR	207	2.20
	WR	48	.52
	RH	16	.17
/S/		6326	
	S	4599	72.69
	C	1067	16.86
	SS	442	6.98
	SC	77	1.21
	X	77	1.21
	ST	33	.52
	PS	19	.30
	Z	6	.14
	SW	4	.06
	SCH	2	.03
/SH/		1537	
	TI	820	53.35
/SH/--Continued			
	SH	398	25.89
	CI	81	5.27
	SSI	51	3.31
	SI	38	2.47
	C	38	2.47
	CH	34	2.21
	T	30	1.95
	S	20	1.30
	SS	9	.58
	SC	6	.39
	SCI	5	.32
	X	3	.23
	CE	2	.13
	SCH	2	.13
/T/		7793	
	T	7528	96.59
	TT	216	2.77
	ED	28	.35
	BT	11	.14
	TW	3	.08
	PT	2	.02
	CT	2	.02
	CHT	2	.02
	TH	1	.01
/T1/		411	
	TH	411	100.00
/T2/		149	
	TH	149	100.00
/V/		1492	
/V/--Continued			
	V	1485	99.54
	LV	4	.26
	F	3	.20
/W/		626	
	W	578	92.35
	U	47	7.50
	OU	1	.15
/Y/		120	
	I	66	55.00
	Y	53	44.17
	E	1	.83
/Z/		995	
	S	640	64.32
	Z	229	23.01
	ES	44	4.42
	X	42	4.22
	ZZ	23	2.33
	SS	13	1.30
	SC	3	.30
	CZ	1	.10
/ZH/		102	
	SI	50	49.01
	S	34	33.33
	G	15	14.70
	Z	2	1.98
	TI	1	.98

---

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

also is spelled with < C >, its second most frequently used graphemic representation, 16.86% of the time. Notice also that < S > is the most frequent spelling for the consonant phoneme < S > and the consonant phoneme /Z/ 72.69% and 64.32%, respectively. Thus, not only do some consonant phonemes fail to attain the 80% criterion, but some of these phonemes also are represented frequently by the same graphemic option, complicating the problem of spelling these phonemes in words which contain them.

Other consonant phonemes falling below the 80% criterion include: the consonant phoneme /Y/, used quite often as a semi-vowel; the consonant phonemes /CH/, /SH/, and /ZH/; and the consonant phoneme /NG/.

What relevant findings accrue from this initial analysis of phoneme-grapheme relationships in the American-English orthography? Generally, it can be stated that these relationships, when other phonological factors are not taken into consideration, inconclusively approximate the alphabetic principle. It is seen that the vowel phonemes present the most difficult spelling problems; conversely, the consonant phonemes have somewhat stable phoneme-grapheme relationships. Obviously, the vowel and consonant phonemes do not equally present similar spelling problems to users of the orthography. Some phonemes are very consistent in their graphemic representations, while others are especially inconsistent, no matter which phonemic alphabet is employed in an analysis of the American-English orthography. In sum, the 52 phonemes of the analysis approximate the alphabetic principle 73.13% of the time. This percentage, however, obscures two very important factors which are relevant to a determination of how nearly the orthography approximates the alphabetic principle. These factors are: (1) the frequency with which each phoneme is used in the lexicon in relationship to all phonemes; that is, the distribution of phonemes throughout the corpus, and (2) the number of graphemic options which represent each phoneme in the corpus.

The Distribution of Phonemes in American-English

It is a misunderstanding of the alphabetic principle to conclude that phonemes are equally distributed throughout the American-English lexicon either with respect to syllabic position or irrespective of this restriction. Obviously, some phonemes are used with greater frequency than others. A moment's reflection will suggest the implications of the distributional properties of phonemes for the problem of spelling. Not only do phonemes have different approximations to the alphabetic principle, but also they are encountered in varying frequency of occurrence in words, irrespective of the frequency of occurrence of the words themselves in running text. Tables 9 and 10 illustrate the distributions of phonemes as they occur in the corpus.

These tables provide the following information:

1. The phoneme classified.
2. Its frequency of occurrence in the corpus.
3. Its percentage of occurrence in relation to the other phonemes.
4. Its rank order in terms of distribution in the corpus for both consonant and vowel phonemes.
5. Its percentage of occurrence in relation to other phonemes of the same classification (i.e., either vowel or consonant).
6. Its rank order in terms of distribution among phonemes of the same classification.

What interpretation do these tables suggest regarding the distribution of phonemes throughout the lexicon? In general, it may be

TABLE 9

FREQUENCY AND PERCENTAGE TABULATIONS OF DISTRIBUTION OF PHONEMES IN AMERICAN-ENGLISH: 22 VOWEL CLASSIFICATION

Phoneme	Total Occurrence in Corpus	% of Occurrence ---- All Phonemes	Rank Order ---- All Phonemes	% of Occurrence ---- Vowels	Rank Order ---- Vowels
A	2248	2.07	17	5.46	7
A2	220	.20	41	.53	19
A3	4340	4.00	9	10.55	3
A5	580	.53	34	1.41	15
E	2538	2.34	15	6.17	6
E2	198	.18	42	.48	20
E3	3646	3.36	11	8.86	4
E5	2170	2.00	18	5.27	8
I	1482	1.36	23	3.60	10
I3	7815	7.20	2	18.99	1
O	2587	2.38	14	6.26	5
OI	149	.14	43	.36	21
OU	406	.37	38	.99	17
O2	767	.71	30	1.86	14
O3	1662	1.53	20	4.04	9
O5	127	.12	44	.31	22
O6	453	.42	36	1.10	16
O7	368	.34	39	.89	18
ə	6013	5.54	6	14.61	2
U	1188	1.09	26	2.89	12
U2	787	.72	29	1.91	13
U3	1410	1.30	24	3.43	11
Totals	41,145	37.90		100.00	

TABLE 10

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF DISTRIBUTION OF PHONEMES IN AMERICAN-ENGLISH: 30 CONSONANTS CLASSIFIED

Pho-neme	Total Occurrence in Corpus	% of Occurrence ---- All Phonemes	Rank Order ---- All Phonemes	% of Occurrence ---- Vowels	Rank Order ---- Vowels
B	2306	2.12	16	3.42	10
CH	564	.52	35	.84	21
D	3703	3.41	10	5.49	7
F	2022	1.86	19	3.00	11
G	1342	1.24	25	1.99	14
H	778	.72	29	1.15	17
HW	89	.08	47	.13	30
J	982	.90	28	1.46	16
K	4714	4.34	8	6.99	6
KS	271	.25	40	.40	23
KW	196	.18	42	.29	24
L	5389	4.96	7	7.99	5
L1	662	.61	31	.98	18
M	3503	3.23	12	5.20	8
M1	97	.09	46	.14	29
N	7662	7.06	4	11.36	3
NG	616	.57	33	.91	20
N1	128	.12	44	.19	26
P	3455	3.17	13	5.12	9
R	9394	8.65	1	13.93	1
S	6328	5.83	5	9.38	4
SH	1537	1.42	21	2.28	12
T	7796	7.18	3	11.56	2
T1	411	.38	37	.61	22
T2	149	.14	43	.22	25
V	1492	1.37	22	2.21	13
W	626	.58	32	.93	19
Y	120	.11	45	.18	27
Z	997	.92	27	1.48	15
ZH	102	.09	46	.15	28
Totals	67,431	62.10		99.98	

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

noticed that the "long" vowel phonemes are more frequently used than are their allophonic variations. This is also observable for the "short" vowel phonemes and their allophonic variations.

Among the consonant phonemes, it is observed that the phonemes /K/, /S/, /F/, and /SH/ occur more frequently than do the phonemes /J/, /Z/, /Y/, /CH/, /ZH/, and /NG/. Each of these ten phonemes falls below the 80% criterion; but note also that when distribution throughout the lexicon is considered, the probability that an individual will encounter the necessity to choose the correct graphemic option for them in different words is much more likely for the first four consonant phonemes than for the latter six phonemes. Thus, the likelihood that an individual will misspell such phonemes is dependent in part upon the number of times they occur in different words.

What does an analysis of distribution of phonemes contribute to an understanding of how closely the orthography approximates the alphabetic principle? For one, it is quite apparent that those phonemes which fall below the 80% criterion will occur as spelling difficulties with varying frequencies in different words. Observe, for example, the two vowel phonemes /ə/ and /06/. These occur in the corpus 6,013 times and 453 times, respectively. In terms of the different words which a writer might need to spell, the odds are much more likely that /ə/ will be misspelled more often than /06/. In sum, it is necessary to include the distribution of phonemes as they occur in different words in the lexicon as one of the factors which must be considered in determining how closely the alphabetic principle is attained in the orthography. It is quite true that a consideration of phonemic distribution in running text would elicit different rank orderings of phonemes. It should be remembered, however, that spelling is essentially a problem of providing correct phoneme-grapheme correspondences in all the words which comprise an individual's speaking and writing vocabularies.

The Number of Graphemic Options Required to Bring Phonemes to an 80% Criterion

Another relevant factor in analyzing the American-English orthography is the number of graphemic options which legitimately could represent each phoneme. In languages which very nearly attain the alphabetic principle, each phoneme would have only one graphemic representation. A few phonemes in American-English, it has been observed, do attain this degree of consistency, or very nearly so. Other phonemes are represented by several graphemic options and with varying frequencies of occurrence.

This matter is especially relevant to a discussion of phoneme-grapheme relationships. Note, for example, in Table 7, page 77, that the frequently used vowel phoneme /I3/ in a 52 phoneme classification has 22 different graphemic options which can represent it. But also observe that, of the 22 graphemic options, < I > represents /I3/ 68.40% of the time and that < Y > represents it 23.04% of the time. Together these two graphemic options represent /I3/ 91.44% of the time in the corpus, while the other graphemic options represent this phoneme less than 9% of the time--and in decreasing frequency of occurrence.

To illustrate this point further, Tables 11 and 12 present a listing of the number of graphemic options required to bring each phoneme falling below the 80% criterion to or above this standard, irrespective of syllabic position or syllabic stress. These tables provide the following information:

1. Which phonemes are below the 80% criterion.

TABLE 11

NUMBER OF GRAPHEMIC OPTIONS REQUIRED TO INCREASE SELECTED PHONEMES BEYOND 80% CRITERION: 22 VOWEL CLASSIFICATION

Phoneme	No. of Options Representing Phoneme	Options Used to Exceed 80% Criterion	% of Occurrence of Options	No. of Options Used to Exceed 80% Criterion	% of Occurrence of Combined Options
/A/	16	A	44.57		
		A-E	35.14		
		AI	9.25	3	88.96
/A2/	9	A	29.09		
		A-E	22.72		
		AI	20.90		
		E-E	15.45	4	88.16
/E/	16	E	69.54		
		EE	9.81		
		EA	9.65	3	89.00
/E2/	8	E	32.32		
		EA	24.74		
		EE	18.18		
		E-E	13.63	4	88.87
/E5/	9	E	76.77		
		O	12.35	2	89.12
/I/	14	I-E	37.44		
		I	37.38		
		Y	14.23	3	89.05
/I3/	22	I	68.40		
		Y	23.04	2	91.44
/O/	16	O	72.51		
		O-E	14.30	2	86.81
/OI/	4	OI	61.74		
		OY	32.31	2	94.05
/OU/	5	OU	55.91		
		OW	29.31	2	85.22

TABLE 11--Continued

Phoneme	No. of Options Representing Phoneme	Options Used to Exceed 80% Criterion	% of Occurrence of Options	No. of Options Used to Exceed 80% Criterion	% of Occurrence of Combined Options
/02/	12	O	40.67		
		A	21.51		
		AU	19.03	3	81.21
/06/	16	OO	38.18		
		U	20.52		
		O	8.16		
		U-E	7.50		
		OU	6.40	5	80.76
/07/	6	U	54.34		
		OO	30.97	2	85.31
/ə/	22	O	26.79		
		A	23.91		
		I	22.40		
		E	12.68	4	85.78
/U/	11	U	68.51		
		U-E	21.54	2	90.05
/U2/	14	E	39.77		
		U	25.79		
		I	13.21		
		O	6.73	4	85.50

TABLE 12

NUMBER OF GRAPHEMIC OPTIONS REQUIRED TO INCREASE SELECTED PHONEMES BEYOND 80% CRITERION: 30 CONSONANT CLASSIFICATION

Phoneme	No. of Options Representing Phoneme	Options Used to Exceed 80% Criterion	% of Occurrence of Options	No. of Options Used to Exceed 80% Criterion	% of Occurrence of Combined Options
/CH/	5	CH	55.49 31.02	2	86.51
/F/	6	F PH	78.25 12.02	2	90.27
/J/	8	G J	65.88 22.24	2	88.12
/K/	13	C K	73.25 12.75	2	86.00
/NG/	3	NG N	58.86 40.81	2	99.67
/N1/	3	EN ON	65.62 32.04	2	97.66
/S/	10	S C	72.69 16.86	2	89.55
/SH/	15	TI SH CI	53.25 25.89 5.27	3	84.51
/Y/	3	I Y	55.00 44.17	2	99.17
/Z/	8	S Z	64.32 23.01	2	87.33
/ZH/	5	SI S	49.01 33.33	2	82.34

2. The number of graphemic options which can represent each phoneme.
3. The graphemic options which, when combined, represent each phoneme over 80% of the time.
4. The percentage of occurrence of each of these graphemic options.
5. The total of the least number of graphemic options which will bring the phoneme above the 80% criterion.
6. The combined percentage tabulations of these combined graphemic options.

It is obvious, of course, that an individual who wishes to spell correctly all the words in his speaking and writing vocabulary cannot dismiss the fact that all the graphemic options for a particular phoneme will be used at some time or other. The task involved in building spelling ability is to be able to associate the appropriate graphemic options for a phoneme in any word, regardless of the frequency of occurrence of the phoneme, or the frequency with which it is represented by a particular graphemic option.

Yet, the fact that a relatively few graphemic options for each phoneme are used over 80% of the time (none requiring more than five options to exceed the criterion) suggests that the rarely used graphemic options present special spelling problems and should be considered as such in terms of basing spelling instruction upon approximations of the orthography to the alphabetic principle. At the least, such frequently occurring phonemes as /I3/ (the next most used phoneme in the lexicon after /R/) can be seen to be spelled predominantly with two graphemic options the vast majority of the time without accounting for positional or stress factors, an indication that a study of these factors may help in better understanding the alphabetic principle underlying the American-English orthography.

Phoneme-Grapheme Correspondences by Position in Syllables

Earlier discussion pointed out the relationships among phonemes and graphemes in terms of their positional occurrences in syllables and monosyllabic words. It is evident that syllables are sequences of phonemes (or graphemes in writing) arranged in particular permissible patterns in American-English. The exact position where a phoneme or grapheme occurs in syllables has considerable relevance to an analysis of the orthography.

It has been seen that the frequency with which a particular phoneme-grapheme correspondence occurs is indicative in part of how closely the orthography approximates the alphabetic principle. It is consequently of additional interest to determine whether positional restrictions have a bearing on the frequency with which certain graphemic options will represent the various phonemes.

The tables that follow contain sufficient information to occupy the time of an investigator far beyond his immediate interests. Consequently, a number of potentially important relationships will not be interpreted. The major conclusions which will be drawn are directed toward a determination of ways in which phoneme-grapheme correspondences can be seen to exceed the 80% criterion.

Of particular interest are those phonemes whose graphemic representations elicit no single graphemic option occurring above the 80% criterion. For the phonemes having only a single graphemic option or one which is used predominantly in the lexicon, positional restrictions are not quite so relevant. But, consider the possible utility of positional effects upon phoneme-grapheme correspondences as typified by the consonant phoneme /NG/ and as tabulated in Table 8, pages 80-81.

This phoneme occurs 615 times in the corpus, with its two most frequently used graphemic options being < NG > and < N >. The option < NG > represents /NG/ 58.86% of the time and the option < N > represents this phoneme 40.81% of the time. Upon the basis of this information, it might be concluded that the speller's odds of choosing the correct graphemic option for /NG/ are limited. An examination of the word listings accompanying the tables of frequency and percentage tabulations--see Appendices and --however, elicits that < NG > only represents /NG/ in word final or syllable final position, with but five exceptions: length, lengthwise, amongst, tongs, and strength, where it is treated as a medially occurring phoneme-grapheme correspondence. Concomitantly, < N > is found to occur primarily in either medial or final position in syllables. Clearly, the odds that < NG > will occur in medial position are insignificant, while an examination of the frequency and percentage tabulations of these two options in final position in syllables shows the preponderance of < NG > in that environment.

Such information does not necessarily suggest that learning about positional influences upon phoneme-grapheme correspondences will solve all the speller's problems. These data do suggest, however, that the environment in which some phonemes occur tends to restrict the choice of graphemic options which may represent them.

22 vowel classification. Table 13 provides the frequency and percentage tabulations of vowel phonemes by initial, medial, and final positions in syllables. This table, and the others to follow, contains the following information:

1. The phoneme classified.
2. The frequency of occurrence of the phoneme in the corpus.

TABLE 13

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN SYLLABLES: 22 VOWEL CLASSIFICATION

Pho-neme	Total Occur-rence	Gra-pheme	Initial		Medial		Final	
			f	%	f	%	f	%
/A/	2248	A	11	1.09	131	13.07	860	85.82
		A-E	121	15.31	669	84.68	0	.00
		AI	5	2.40	187	89.90	16	7.69
		AY	0	.00	3	2.29	128	97.70
		AI-E	1	5.55	17	94.44	0	.00
		EIGH	1	5.55	7	38.88	10	55.55
		E	0	.00	0	.00	16	100.00
		EA	0	.00	13	92.85	1	7.14
		EI	0	.00	11	78.57	3	21.42
		EY	0	.00	0	.00	14	100.00
		ET	0	.00	0	.00	9	100.00
		E-E	0	.00	2	33.33	4	66.66
		AIGH	0	.00	4	100.00	0	.00
		EI-E	0	.00	2	100.00	0	.00
		AU-E	0	.00	1	100.00	0	.00
		AY-E	0	.00	0	.00	1	100.00
/A2/	220	A	1	1.56	63	98.43	0	.00
		A-E	0	.00	50	100.00	0	.00
		AI	9	19.56	37	80.43	0	.00
		E-E	1	2.94	33	97.05	0	.00
		EA	0	.00	13	100.00	0	.00
		EI	3	60.00	2	40.00	0	.00
		E	0	.00	4	100.00	0	.00
		AI-E	3	100.00	0	.00	0	.00
		EY	1	100.00	0	.00	0	.00
/A3/	4340	A	1009	24.06	2879	68.67	304	7.25
		A-E	51	34.69	96	65.30	0	.00
		AI	0	.00	1	100.00	0	.00
/A5/	580	A	88	16.98	372	71.81	58	11.19
		A-E	1	3.22	30	96.77	0	.00
		EA	2	11.11	16	88.88	0	.00
		E	3	60.00	2	40.00	0	.00
		AH	0	.00	1	25.00	3	75.00
		AU	1	25.00	3	75.00	0	.00

TABLE 13--Continued

Pho-neme	Total Occur-rence	Gra-pheme	Initial		Medial		Final	
			f	%	f	%	f	%
/E/	2538							
		E	1	.05	24	1.35	1740	98.58
		EE	1	.40	1661	66.66	82	32.93
		EA	17	6.93	175	71.42	53	21.63
		E-E	3	4.83	59	95.16	0	.00
		I-E	1	2.27	43	97.72	0	.00
		I	1	2.63	14	36.84	23	60.52
		IE	0	.00	33	100.00	0	.00
		EA-E	1	3.33	29	96,66	0	.00
		IE-E	0	.00	23	100.00	0	.00
		EI	0	.00	11	68.75	5	31.25
		EE-E	0	.00	9	100.00	0	.00
		EI-E	0	.00	6	100.00	0	.00
		EY	0	.00	0	.00	6	100.00
		AE	0	.00	0	.00	5	100.00
		OE	0	.00	0	.00	5	100.00
		EO	0	.00	0	.00	2	100.00
/E2/	198							
		E	1	1.56	63	98.43	0	.00
		EA	4	8.16	45	91.83	0	.00
		EE	2	5.55	34	94.44	0	.00
		E-E	0	.00	27	100.00	0	.00
		IE	1	7.14	13	92.85	0	.00
		I	0	.00	3	100.00	0	.00
		IE-E	0	.00	3	100.00	0	.00
		EI	0	.00	2	100.00	0	.00
/E3/	3646							
		E	749	22.58	2523	76.08	44	1.32
		EA	0	.00	139	100.00	0	.00
		A	32	34.04	62	65.95	0	.00
		E-E	15	18.98	64	81.01	0	.00
		AI	0	.00	4	100.00	0	.00
		IE	0	.00	4	100.00	0	.00
		EO	0	.00	3	100.00	0	.00
		U	0	.00	2	100.00	0	.00
		A-E	1	100.00	0	.00	0	.00
		AE	1	100.00	0	.00	0	.00
		AY	0	.00	1	100.00	0	.00
		EA-E	0	.00	1	100.00	0	.00
		EI	0	.00	1	100.00	0	.00
/E5/	2170							
		E	820	49.21	840	50.42	6	.36
		O	70	26.11	198	73.88	0	.00
		A	52	30.95	116	69.04	0	.00
		U	8	25.80	23	74.19	0	.00

TABLE 13--Continued

Phoneme	Total Occurrence	Grapheme	Initial f	Initial %	Medial f	Medial %	Final f	Final %
/E5/--Continued								
		U-E	8	34.78	15	65.21	0	.00
		I	1	12.50	7	87.50	0	.00
		Y	2	50.00	2	50.00	0	.00
		E-E	1	100.00	0	.00	0	.00
		OU	1	100.00	0	.00	0	.00
/I/	1482							
		I-E	52	9.36	503	90.63	0	.00
		I	2	.36	157	28.33	395	71.29
		Y	0	.00	2	.94	209	99.05
		IGH	0	.00	78	88.63	10	11.36
		IE	0	.00	14	53.84	12	46.15
		Y-E	0	.00	19	82.60	4	17.39
		EY-E	0	.00	0	.00	7	100.00
		EI	0	.00	5	83.33	1	16.66
		AY	0	.00	0	.00	3	100.00
		EIGH	0	.00	3	100.00	0	.00
		UY	0	.00	0	.00	3	100.00
		AI-E	1	100.00	0	.00	0	.00
		EY	0	.00	0	.00	1	100.00
		OY	0	.00	0	.00	1	100.00
/I3/	7815							
		I	1535	28.71	2772	51.85	1030	19.43
		Y	5	.27	149	8.27	1647	91.44
		I-E	44	12.97	295	87.02	0	.00
		A-E	108	57.75	79	42.24	0	.00
		EY	0	.00	0	.00	40	100.00
		IE	5	17.24	8	27.58	16	55.17
		UI	0	.00	15	93.75	1	6.25
		AI	1	6.66	14	93.33	0	.00
		EI	8	72.72	3	27.27	0	.00
		EE	0	.00	2	33.33	4	66.66
		E	0	.00	2	40.00	3	60.00
		A	4	100.00	0	.00	0	.00
		IA-E	3	100.00	0	.00	0	.00
		U	0	.00	3	100.00	0	.00
		U-E	3	100.00	0	.00	0	.00
		AY	0	.00	0	.00	1	100.00
		E-E	0	.00	1	100.00	0	.00
		EA	0	.00	0	.00	1	100.00
		EA-E	1	100.00	0	.00	0	.00
		IE-E	0	.00	1	100.00	0	.00
		O	0	.00	1	100.00	0	.00
		Y-E	0	.00	1	100.00	0	.00

TABLE 13--Continued

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/O/	2587							
		O	5	.26	242	12.89	1629	86.83
		O-E	11	2.97	359	97.02	0	.00
		OA	8	6.34	116	92.06	2	1.58
		OW	7	5.64	14	11.29	103	83.06
		OU	0	.00	29	100.00	0	.00
		OE	0	.00	0	.00	13	100.00
		OU-E	0	.00	10	100.00	0	.00
		OO	0	.00	9	100.00	0	.00
		OUGH	0	.00	0	.00	8	100.00
		EAU	0	.00	0	.00	6	100.00
		OH	1	25.00	0	.00	3	75.00
		AU	0	.00	0	.00	3	100.00
		EW	0	.00	0	.00	3	100.00
		OA-E	0	.00	3	100.00	0	.00
		AU-E	0	.00	2	100.00	0	.00
		OW-E	0	.00	0	.00	1	100.00
/OI/	149							
		OI	8	8.69	71	77.17	13	14.13
		OY	0	.00	0	.00	48	100.00
		OI-E	0	.00	8	100.00	0	.00
		OY-E	0	.00	1	100.00	0	.00
/OU/	406							
		OU	69	30.39	153	67.40	5	2.20
		OW	2	1.68	45	37.81	72	60.50
		OU-E	1	1.85	53	98.14	0	.00
		OUGH	0	.00	1	25.00	3	75.00
		OW-E	0	.00	2	100.00	0	.00
/O2/	767							
		O	46	14.74	266	85.25	0	.00
		A	26	15.75	125	75.75	14	8.48
		AU	9	6.16	57	39.04	80	54.79
		AW	2	2.66	22	29.33	51	68.00
		O-E	0	.00	17	100.00	0	.00
		OUGH	1	6.66	14	93.33	0	.00
		AUGH	1	8.33	6	50.00	5	41.66
		OA	0	.00	9	100.00	0	.00
		AU-E	0	.00	9	100.00	0	.00
		A-E	0	.00	3	100.00	0	.00
		AO	0	.00	2	100.00	0	.00
		AW-E	0	.00	0	.00	2	100.00
/O3/	1662							
		O	214	13.73	1343	86.20	1	.06
		A	0	.00	80	100.00	0	.00

TABLE 13--Continued

Pho-neme	Total Occur-rence	Gra-pheme	Initial		Medial		Final	
			f	%	f	%	f	%
/03/--Continued								
		O-E	0	.00	20	100.00	0	.00
		OW	0	.00	4	100.00	0	.00
/05/	127							
		O	11	8.94	112	91.05	0	.00
		O-E	0	.00	4	100.00	0	.00
/06/	453							
		OO	0	.00	148	85.54	25	14.45
		U	0	.00	11	11.82	82	88.17
		O	0	.00	16	43.24	21	56.75
		U-E	0	.00	34	100.00	0	.00
		OU	0	.00	9	31.03	20	68.96
		EW	0	.00	2	9.09	20	90.90
		UE	0	.00	0	.00	16	100.00
		OO-E	1	8.33	11	91.66	0	.00
		O-E	0	.00	12	100.00	0	.00
		UI	0	.00	6	100.00	0	.00
		EU	0	.00	1	25.00	3	75.00
		OE	0	.00	0	.00	4	100.00
		UI-E	0	.00	4	100.00	0	.00
		OU-E	0	.00	3	100.00	0	.00
		OUGH	0	.00	0	.00	2	100.00
		UO	0	.00	0	.00	2	100.00
/07/	368							
		U	0	.00	164	82.00	36	18.00
		OO	0	.00	114	100.00	0	.00
		OU	0	.00	21	84.00	4	16.00
		O	0	.00	11	64.70	6	35.29
		U-E	0	.00	11	100.00	0	.00
		EU	0	.00	1	100.00	0	.00
/ə/	6013							
		O	321	19.92	1176	72.99	114	7.07
		A	6	.41	13	.90	1419	98.67
		I	6	.44	9	.66	1332	98.88
		E	228	29.88	420	55.04	115	15.07
		OU	191	56.84	144	42.85	1	.29
		U	90	30.30	165	55.55	42	14.14
		E-E	37	36.63	64	63.36	0	.00
		Y	2	8.69	1	4.34	20	86.95
		IE	13	59.09	7	31.81	2	9.09
		O-E	1	4.76	20	95.23	0	.00
		EO	5	50.00	5	50.00	0	.00
		AI	3	33.33	2	22.22	4	44.44
		EOU	0	.00	8	100.00	0	.00

TABLE 13--Continued

Pho-neme	Total Occur-rence	Gra-pheme	Initial f	Initial %	Medial f	Medial %	Final f	Final %
/ə/--Continued								
		I-E	2	33.33	4	66.66	0	.00
		IE-E	1	25.00	3	75.00	0	.00
		U-E	1	25.00	0	.00	3	75.00
		A-E	1	33.33	2	66.66	0	.00
		EA	2	66.66	1	33.33	0	.00
		EI	1	50.00	0	.00	1	50.00
		IA	0	.00	0	.00	2	100.00
		OI-E	0	.00	2	100.00	0	.00
		AU	0	.00	0	.00	1	100.00
/U/	1188							
		U	9	1.10	35	4.29	770	94.59
		U-E	31	12.10	225	87.89	0	.00
		EW	0	.00	7	18.42	31	81.57
		EU	0	.00	2	7.14	26	92.85
		UE	1	3.70	0	.00	26	96.29
		UI	0	.00	7	87.50	1	12.50
		IEW	0	.00	0	.00	6	100.00
		EAU	0	.00	0	.00	5	100.00
		IEU	0	.00	0	.00	4	100.00
		EU-E	0	.00	1	100.00	0	.00
		EW-E	0	.00	0	.00	1	100.00
/U2/	787							
		E	10	3.19	303	96.80	0	.00
		U	8	3.94	195	96.05	0	.00
		I	2	1.92	102	98.07	0	.00
		O	0	.00	53	100.00	0	.00
		E-E	2	4.87	39	95.12	0	.00
		EA	15	51.72	14	48.27	0	.00
		OU	0	.00	21	100.00	0	.00
		U-E	1	10.00	9	90.00	0	.00
		EU	0	.00	6	100.00	0	.00
		EA-E	0	.00	2	100.00	0	.00
		Y	0	.00	2	100.00	0	.00
		I-E	0	.00	1	100.00	0	.00
		O-E	0	.00	1	100.00	0	.00
		OU-E	0	.00	1	100.00	0	.00
/U3/	1410							
		U	270	22.27	940	77.55	2	.16
		O	5	4.46	107	95.53	0	.00
		OU	0	.00	30	100.00	0	.00
		O-E	0	.00	26	100.00	0	.00
		U-E	0	.00	23	100.00	0	.00
		OO	0	.00	7	100.00	0	.00

---

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

3. The different graphemic options representing the phoneme.
4. The frequency and percentage tabulation of the occurrence of each graphemic option for the phoneme in initial position in syllables.
5. The frequency and percentage tabulation of the occurrence of each graphemic option for the phoneme in medial position in syllables.
6. The frequency and percentage tabulation of the occurrence of each graphemic option for the phoneme in final position in syllables.

/A/: In Table 13, note first that < A >, this phoneme's most frequently used graphemic option, occurs 85.82% of the time in final position in syllables. < A-E >, on the other hand, occurs in medial position 84.68% of the time it is used. Since such graphemic options as < A-E > are comprised of an internal vowel phoneme and the grapheme < E > following an intervening consonant phoneme, such options are usually considered to be in either initial or medial positions in syllables. (On occasion, graphemic options of the type < AY-E > will be listed as finally occurring options in syllables. < AY-E > is actually the spelling of the word aye, which is a single vowel phoneme represented by three graphemes. In these cases, the computer classifies the vowel phoneme as occurring in final position in a syllable and therefore its entire spelling is listed in that position as well. This phenomenon will be found to occur for certain other phonemes throughout the tables to follow.)

Observe that < E >, < ET >, and < EY > occur only in final position and the infrequently used options < AIGH >, < EI-E >, and < AU-E > occur only medially.

/A2/: This phoneme primarily occurs in medial position before the phoneme /R/. Notice that the rarely used graphemic options < AI-E > and < EY > occur only in initial position in syllables.

/E/: The graphemic option < E >, when used, occurs in final position 98.58% of the time. Notice that < IE >, < IE-E >, < EE-E >, and < EI-E > occur only in medial position, while < EY >, < AE >, < OE >, and < EO > occur only in final position in syllables.

/E2/: This phoneme occurs before /R/. Notice that < E-E >, < I >, < IE-E >, and < EI > occur only in medial position in syllables.

/E5/: This phoneme primarily occurs before /R/ and thus has few occurrences in final position in syllables. Observe that < I >, when used, occurs 87.50% of the time in medial position.

/I/: This phoneme occurs mainly in medial and final positions in syllables. The 52 occurrences of < I-E > in initial position are typified by the word ice. Observe also that there are several instances in which particular graphemic options occur only in one position in syllables, these being < EY-E >, < AY >, < EIGH >, < UY >, < AI-E >, < EY >, and < OY >.

/I3/: Notice that < I > is rather evenly distributed among the three positions in syllables, yet < Y > occurs 91.44% of the time in final position. Note also that < EY > occurs only in final position, and that the remaining infrequently used graphemic options generally are confined to particular positions in syllables.

/O/: Observe that this phoneme's most frequently used graphemic option, < O >, when used, occurs in final position in syllables 86.83%

of the time, while the spelling < O-E >, when used, occurs in medial position 97.02% of the time. Note also that < OA > is used 92.06% of the time in medial position, while < OW > is used 83.06% of the time in final position. Observe as well that < OU >, < OU-E >, < OO >, < OA-E >, and < AU-E > occur only in medial position in syllables and that < OUGH >, < EAU >, < AU >, < EW >, and < OW-E > (the word owe) occur only in final position.

/OI/: The graphemic option < OY > appears to be used only in final position in syllables. This is not entirely accurate; one word, oyster, uses this option in initial position. Its omission from the listing is due to one of the infrequent coding errors that occurred when the data were transcribed onto data processing cards and thus appears on the "true error" list. Note, however, that < OI > represents /OI/ predominantly in medial position in syllables.

/OU/: Observe that < OU-E >, when used, occurs 98.14% of the time in medial position. Otherwise, an anlysis of positional influences on phoneme-grapheme correspondences does not provide any useful evidence concerning its most frequently used options < OU > and < OW >, although the rarely used options < OU-E > and OW-E > are confined to medial position in syllables.

/O2/: This phoneme's most frequent spelling, < O >, when used, occurs in medial position 85.25% of the time. Graphemic options that occur only in medial position are < O-E >, < OA >, < AU-E >, < A-E >, and < AO >. The infrequently used graphemic option < AW-E > occurs only in final position and is the spelling of the word awe.

/O6/: This phoneme has many infrequently used phoneme-grapheme correspondences. Having 16 different graphemic options, five of them are required to bring this particular phoneme above the 80% criterion. Observe, however, that an analysis of positional influences somewhat clarifies a seemingly confusing pattern of phoneme-grapheme relationships. When used, < OO > occurs 85.54% of the time in medial position in syllables, while < U >, when used, occurs 88.17% of the time in final position. Note also that < U-E >, < O-E >, < UI >, < UI-E >, and < OU-E > occur only in medial position, while < UE >, < OUGH >, and < UO > occur only in final position. The single occurrence of < OO-E > in initial position in a syllable is confined to the word ooze.

/O7/: This particular phoneme occurs primarily in medial position in syllables.

/ə/: This phoneme, used almost entirely in unaccented syllables, is responsible for much of the misunderstanding regarding the alphabetic nature of the orthography. Note its frequent use and its many graphemic representations. But observe also that < O >, when used, although not attaining the 80% criterion, occurs 72.99% of the time in medial position, while the graphemic options < A > and < I > occur almost always in final position in syllables. In addition, 12 of the graphemic options of /ə/ occur less than ten times each in over 17,000 different words. Over half of the graphemic options of /ə/ are infrequent spellings and should ordinarily be taught as such.

/U/: This phoneme occurs mostly in final position in syllables, although < U >, when used, occurs 87.89% of the time in medial position. Observe also the one occurrence of < EW-E >, which is the spelling of the word ewe.

/U2/: This phoneme occurs before /R/ and is mainly used in medial position in syllables. It is used in initial position in these few cases: < E >, < U >, < I >, < E-E >, < EA >, and < U-E >.

30 consonant classification. As was discussed previously, numerous occurrences of consonant phonemes in medial position in syllables will appear in the data appearing in Table 14. Aside from a few instances in which these tabulations provide pertinent information, the interpretation to follow will be restricted to consonant phoneme-grapheme correspondences in initial and final positions in syllables.[1]

/CH/: Observe that the infrequently used graphemic option < TCH > occurs only in final position, and that < TI > and < C > occur only in initial position in syllables.

/F/: Note that < F >, when used, occurs in initial position 87.40% of the time, while < CH > and < FT > occur only in final position in syllables.

/J/: Note that < J >, when used, occurs 98.16% of the time in initial position; observe also that < DG > and < GG > occur only in final position and that < DI > occurs only in initial position in syllables.

/K/: Note that < CK >, when used, occurs in final position 99.66% of the time, and that many of the lesser used options are mainly confined to particular positions in syllables.

/NG/: Note that the graphemic option < NG > occurs in final position 98.62% of the time, while < N > is distributed rather evenly between medial and final positions. The medial occurring < N > is found in such words as sink, drank, bunk, and monk.

/N1/: This phoneme is a syllabic consonant and is used to end such words as cousin and cotton. It is therefore listed as having only

---

[1] It may help the reader to interpret the text to follow concerning consonant phoneme-grapheme correspondences if he refers to Table 5, pages 61-65, for words illustrating each of the cases under discussion.

TABLE 14

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN SYLLABLES: 30 CONSONANT CLASSIFICATION

Phoneme	Total Occurrence	Grapheme	Initial f	Initial %	Medial f	Medial %	Final f	Final %
/B/	2303							
		B	1847	82.49	1	.04	391	17.47
		BB	5	7.93	0	.00	58	92.07
		PB	1	100.00	0	.00	0	.00
/CH/	564							
		CH	177	56.54	2	.63	134	42.83
		T	127	72.57	0	.00	48	27.43
		TCH	0	.00	0	.00	61	100.00
		TI	13	100.00	0	.00	0	.00
		C	2	100.00	0	.00	0	.00
/D/	3691							
		D	2097	58.07	29	.80	1485	41.13
		DD	7	9.45	1	1.35	66	89.20
		LD	2	33.33	0	.00	4	66.67
/F/	2019							
		F	1381	87.40	43	2.72	156	9.88
		PH	182	75.20	9	3.71	51	21.09
		FF	62	35.02	0	.00	115	64.98
		LF	5	55.55	0	.00	4	44.45
		GH	0	.00	0	.00	8	100.00
		FT	0	.00	0	.00	3	100.00
/G/	1338							
		G	871	73.93	1	.08	306	25.99
		GG	5	7.46	0	.00	62	92.54
		X	0	.00	0	.00	42	100.00
		GUE	0	.00	0	.00	21	100.00
		GH	9	90.00	0	.00	1	10.00
		TG	1	100.00	0	.00	0	.00
/H/	778							
		H	762	100.00	0	.00	0	.00
		WH	16	100.00	0	.00	0	.00
/HW/	89							
		WH	89	100.00	0	.00	0	.00
/H9/	41							
		H	21	51.21	11	26.82	9	21.97

TABLE 14--Continued

Pho-neme	Total Occur-rence	Gra-pheme	Initial		Medial		Final	
			f	%	f	%	f	%
/J/	982							
		G	340	52.55	0	.00	307	47.45
		J	214	98.16	0	.00	4	1.84
		DG	0	.00	0	.00	51	100.00
		D	7	21.87	0	.00	25	78.13
		DJ	13	81.25	0	.00	3	18.75
		GI	11	78.57	0	.00	3	21.43
		GG	0	.00	0	.00	2	100.00
		DI	2	100.00	0	.00	0	.00
/K/	4712							
		C	2287	66.25	286	8.28	879	25.47
		K	162	26.95	47	7.82	392	65.23
		CK	0	.00	1	.34	289	99.66
		CH	96	67.60	13	9.15	33	23.25
		X	0	.00	0	.00	80	100.00
		CC	46	60.52	0	.00	30	39.48
		QU	8	29.62	0	.00	19	70.38
		Q	2	10.00	0	.00	18	90.00
		LK	0	.00	0	.00	14	100.00
		CQ	0	.00	0	.00	3	100.00
		KH	3	100.00	0	.00	0	.00
		SC	3	100.00	0	.00	0	.00
		CCH	0	.00	0	.00	1	100.00
/KS/	271							
		X	0	.00	8	3.26	237	96.74
		CS	0	.00	0	.00	26	100.00
/KW/	196							
		QU	165	86.39	26	13.61	0	.00
		CQU	5	100.00	0	.00	0	.00
/L/	5389							
		L	1769	36.14	1108	22.63	2017	41.23
		LL	104	21.26	1	..20	384	78.54
		SL	3	60.00	0	.00	2	40.00
		LN	0	.00	0	.00	1	100.00
/Ll/	651							
		L	0	.00	1	.16	619	99.84
		EL	0	.00	0	.00	19	100.00
		IL	0	.00	0	.00	7	100.00
		AL	0	.00	0	.00	4	100.00
		OL	0	.00	0	.00	1	100.00

TABLE 14--Continued

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/M/	3501							
		M	1874	56.75	119	3.60	1309	39.65
		MM	61	43.57	0	.00	79	56.43
		MB	0	.00	0	.00	27	100.00
		LM	2	11.76	2	11.76	13	76.48
		GM	5	62.50	0	.00	3	37.50
		MN	0	.00	0	.00	7	100.00
/M1/	97							
		M	0	.00	0	.00	97	100.00
/N/	7656							
		N	1223	16.41	1368	18.35	4861	65.24
		NN	30	23.62	0	.00	97	76.38
		KN	41	100.00	0	.00	0	.00
		GN	5	15.62	0	.00	27	84.38
		PN	3	100.00	0	.00	0	.00
		MN	1	100.00	0	.00	0	.00
/NG/	615							
		NG	0	.00	5	1.38	357	98.62
		N	0	.00	129	51.39	122	48.61
		NGUE	0	.00	0	.00	2	100.00
/N1/	128							
		EN	0	.00	0	.00	84	100.00
		ON	0	.00	0	.00	41	100.00
		IN	0	.00	0	.00	3	100.00
/P/	3449							
		P	2302	69.84	277	8.40	717	21.76
		PP	66	43.13	1	.65	86	56.22
/R/	9390							
		R	1648	18.07	3175	34.81	4296	47.12
		RR	50	24.15	0	.00	157	75.85
		WR	48	100.00	0	.00	0	.00
		RH	16	100.00	0	.00	0	.00
/S/	6326							
		S	2634	57.27	373	8.11	1592	34.62
		C	537	50.32	10	.93	520	48.75
		SS	70	15.83	2	.45	370	83.72
		SC	39	50.64	0	.00	38	49.36
		X	77	100.00	0	.00	0	.00
		ST	0	.00	0	.00	33	100.00
		PS	19	100.00	0	.00	0	.00

TABLE 14--Continued

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/S/--Continued								
		Z	2	33.33	0	.00	4	66.67
		SW	4	100.00	0	.00	0	.00
		SCH	2	100.00	0	.00	0	.00
/SH/	1537							
		TI	729	88.90	0	.00	91	11.10
		SH	200	50.25	4	1.00	194	48.75
		CI	39	48.14	0	.00	42	51.86
		SSI	0	.00	0	.00	51	100.00
		SI	38	100.00	0	.00	0	.00
		C	22	57.89	0	.00	16	42.11
		CH	29	85.29	0	.00	5	14.71
		T	24	80.00	0	.00	6	20.00
		S	19	95.00	0	.00	1	5.00
		SS	5	55.55	0	.00	4	44.45
		SC	4	66.66	0	.00	2	33.33
		SCI	4	80.00	0	.00	1	20.00
		X	3	100.00	0	.00	0	.00
		CE	2	100.00	0	.00	0	.00
		SCH	2	100.00	0	.00	0	.00
/T/	7793							
		T	3722	49.44	526	6.98	3280	43.58
		TT	28	12.96	0	.00	188	87.04
		ED	0	.00	0	.00	28	100.00
		BT	0	.00	0	.00	11	100.00
		TW	3	100.00	0	.00	0	.00
		PT	1	50.00	0	.00	1	50.00
		CT	0	.00	0	.00	2	100.00
		CHT	1	50.00	0	.00	1	50.00
		TH	1	100.00	0	.00	0	.00
/T1/	411							
		TH	226	54.98	0	.00	185	45.02
/T2/	149							
		TH	61	40.93	4	2.68	84	56.39
/V/	1492							
		V	838	56.43	16	1.07	631	42.50
		LV	0	.00	2	50.00	2	50.00
		F	0	.00	0	.00	3	100.00

TABLE 14--Continued

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/W/	626							
		W	494	85.47	84	14.53	0	.00
		U	22	46.81	25	53.19	0	.00
		OU	1	100.00	0	.00	0	.00
/Y/	120							
		I	64	96.97	2	3.03	0	.00
		Y	53	100.00	0	.00	0	.00
		E	1	100.00	0	.00	0	.00
/Z/	995							
		S	168	26.25	96	15.00	376	58.75
		Z	84	36.68	0	.00	145	63.32
		ES	0	.00	0	.00	44	100.00
		X	42	100.00	0	.00	0	.00
		ZZ	1	4.34	0	.00	22	95.66
		SS	11	84.61	0	.00	2	15.39
		SC	3	100.00	0	.00	0	.00
		CZ	1	100.00	0	.00	0	.00
/ZH/	102							
		SI	35	70.00	0	.00	15	30.00
		S	19	55.88	0	.00	15	44.12
		G	3	20.00	0	.00	12	80.00
		Z	1	50.00	0	.00	1	50.00
		TI	1	100.00	0	.00	0	.00

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

final occurring phoneme-grapheme correspondences in syllables.

/S/: Observe that < I >, < PS >, < SW >, and < SCH >, when used, are each found only in initial position, while < ST > occurs only in final position in syllables. Observe also that < SS >, when used, occurs 83.72% of the time in final syllables.

/SH/: Note that its most frequent spelling, < TI >, when used, occurs 88.90% of the time in initial position, as in the second syllable of nation, and that < SSI > occurs only in final position. Notice also that the remaining less frequently used options are confined predominantly to initial position in syllables.

/Y/: This consonant phoneme is confined primarily to initial position in syllables, having only two occurrences of < I > in medial position.

/Z/: Note that the less frequently used graphemic options are mainly confined to a particular position in syllables, with < ES > occurring only in final position, and < X >, < SC >, and < CZ > occurring only in initial position.

/ZH/: Observe that < G >, when used, occurs 80.00% of the time in final position in syllables, and that the two unique options < Z > and < TI > occur in initial position.

In summarizing the effects of position in syllables upon the selection of graphemic options for phonemes, it is seen that positional factors do not equally influence all phonemes. In some instances, certain graphemic options are relegated to a single position, while in others, occurrences of phoneme-grapheme correspondences are distributed almost randomly throughout syllables and words.

Of most interest are the effects of positional factors on the graphemic options for phonemes which fall below the 80% criterion. It is noted that, for the vowel phonemes, these influences are most noticeable in medial and final positions in syllables. Nearly all the vowel phonemes below the 80% criterion in either phonemic classification have one or more graphemic options that will occur 80% or more of the time in a particular position. In a few cases, these graphemic options are among those options which most frequently represent particular phonemes.

Observe, for example, the increased restrictions placed upon certain frequently used options for /A/, /A2/, /E/, /I/, /I3/, /O/, /O1/, /O2/, /O6/, /ə/, /U/, and /U2/ in a 52 phoneme classification. Likewise, notice that the consonant phonemes /F/, /NG/, /NL/, /SH/, and /Y/ are assigned frequently used graphemic options in particular positions in syllables over 80% of the time.

It is not the authors' purpose to examine thoroughly each of these phoneme-grapheme correspondences beyond the conclusion that the occurrences of these correspondences in certain environments in syllables does affect relationships between phonemes and their graphemic representations. A further study of these positional effects, however, conceivably could do much to achieve an ordered description of how closely the orthography approximates the alphabetic principle when certain phonological factors are considered.

Accordingly, it would be of additional interest to determine whether phoneme-grapheme correspondences are not only influenced by where they occur in syllables, but whether the amounts of stress given to syllables tends to further restrict phoneme-grapheme relationships.

Phoneme-Grapheme Correspondences by Position in Stressed and Unstressed Syllables

Three amounts of stress are given in the Merriam-Webster's New Collegiate Dictionary, sixth edition. These stress markers have been classified as primary, secondary, and unaccented, respectively. Primary stress is the stress given to monosyllabic words in isolation and to the syllable in polysyllabic words which is most heavily stressed when pronouncing the word. Secondary stress occurs when a slightly lighter stress is given to a primary stressed syllable, or to a monosyllabic word when affixes are added to it. The so-called unaccented syllable is one which receives the lightest stress in pronouncing a polysyllabic word.

As the amount of accent lessens, vowel phonemes in syllables become more difficult to discriminate. The phoneme /ə/ is the phonemic symbol used to classify vowel phonemes occurring in unaccented syllables and which are pronounced much like the phoneme /U3/ in cup. It is for this reason that /ə/ has many graphemic options; its graphemic representations are mainly a collection of the graphemic options for many different vowel phonemes appearing in unaccented syllables.

As vowels are less able to be discriminated as syllabic stress decreases, some influence will occur on where certain vowel phoneme-grapheme correspondences are used in syllables. The open syllable (i.e., one which ends with a vowel phoneme) usually includes a "long" vowel sound.

Given the assumption that vowel phonemes tend toward shortened pronunciations as syllabic stress decreases and are reduced to /ə/, fewer vowel phoneme-grapheme correspondences will be found in final position in syllables since "short" vowel phonemes usually occur in medial position. It was the purpose of this investigation to determine whether the amount of stress given to syllables in the corpus tended to restrict phoneme-grapheme correspondences to certain positions in these syllables.

22 vowel classification. /A/: Observe, in Table 15, that this phoneme occurs predominantly in primary accented syllables, although the graphemic option < A-E > has numerous occurrences in unaccented syllables.

/A2/: This phoneme is restricted mainly to primary accented syllables.

/E/: This phoneme occurs predominantly in primary accented syllables. Observe, however, that the graphemic option < E > mainly occurs in unaccented syllables.

/E2/: This phoneme occurs predominantly in primary accented syllables.

/E3/: This phoneme occurs only in unaccented syllables, with the exception of its use in the word the.

/I/: It is observed that < I >, when used medially, occurs 89.17% of the time in primary accented syllables; notice also that this phoneme is distributed in medial and final positions in syllables having all levels of stress, with the exceptions of a few occurrences of < I-E >, < I >, and < AI-E > in initial position of primary or unaccented syllables.

/I3/: In initial position, this phoneme occurs mainly in unaccented syllables, with the exception of numerous occurrences of < I > in primary and secondary accented syllables, and in final position only in unaccented syllables.

/O/: This phoneme occurs mainly in primary accented syllables. Observe, however, the numerous occurrences of < O >, < O-E >, and < OW > in unaccented syllables.

/OI/: This phoneme is mainly restricted to occurrences in primary and secondary accented syllables.

/OU/: This phoneme occurs predominantly in primary and secondary accented syllables.

/O2/: This phoneme occurs mainly in primary accented syllables. Note that many of its infrequently used graphemic options occur only in primary accented syllables.

TABLE 15

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN STRESSED AND UNSTRESSED SYLLABLES: 22 VOWEL CLASSIFICATION

Pho-neme	Total Occur-rence	Gra-pheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/A/	2248	A	6	54.54	3	27.27	2	18.18
		A-E	13	10.74	3	2.47	105	86.77
		AI	4	80.00	0	.00	1	20.00
		AY	0	.00	0	.00	0	.00
		AI-E	1	100.00	0	.00	0	.00
		EIGH	1	100.00	0	.00	0	.00
		E	0	.00	0	.00	0	.00
		EA	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
		EY	0	.00	0	.00	0	.00
		ET	0	.00	0	.00	0	.00
		E-E	0	.00	0	.00	0	.00
		AIGH	0	.00	0	.00	0	.00
		EI-E	0	.00	0	.00	0	.00
		AU-E	0	.00	0	.00	0	.00
		AY-E	0	.00	0	.00	0	.00
/A2/	220	A	1	100.00	0	.00	0	.00
		A-E	0	.00	0	.00	0	.00
		AI	9	100.00	0	.00	0	.00
		E-E	1	100.00	0	.00	0	.00
		EA	0	.00	0	.00	0	.00
		EI	3	100.00	0	.00	0	.00
		E	0	.00	0	.00	0	.00
		AI-E	3	100.00	0	.00	0	.00
		EY	1	100.00	0	.00	0	.00
/A3/	4340	A	336	33.30	104	10.30	569	56.39
		A-E	0	.00	0	.00	51	100.00
		AI	0	.00	0	.00	0	.00
/A5/	580	A	60	68.18	8	9.09	20	22.72
		A-E	1	100.00	0	.00	0	.00
		EA	0	.00	2	100.00	0	.00
		E	2	66.66	1	33.33	0	.00
		AH	0	.00	0	.00	0	.00
		AU	1	100.00	0	.00	0	.00

TABLE 15--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
105	80.15	25	19.08	1	.76	778	90.46	71	8.25	11	1.27
315	47.08	31	4.63	323	48.28	0	.00	0	.00	0	.00
168	89.83	14	7.48	5	2.67	14	87.50	0	.00	2	12.50
0	.00	2	66.66	1	33.33	81	63.28	40	31.25	7	5.46
16	94.11	0	.00	1	5.88	0	.00	0	.00	0	.00
4	57.14	3	42.85	0	.00	10	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	5	31.25	1	6.25	10	62.50
7	53.84	6	46.15	0	.00	1	100.00	0	.00	0	.00
11	100.00	0	.00	0	.00	3	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	14	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	6	66.66	0	.00	3	33.33
2	100.00	0	.00	0	.00	3	75.00	0	.00	1	25.00
3	75.00	1	25.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
59	93.65	4	6.34	0	.00	0	.00	0	.00	0	.00
42	84.00	8	16.00	0	.00	0	.00	0	.00	0	.00
31	83.78	4	10.81	2	5.40	0	.00	0	.00	0	.00
7	21.21	8	24.24	18	54.54	0	.00	0	.00	0	.00
11	84.61	2	15.38	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	25.00	0	.00	3	75.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
1808	62.79	237	8.23	834	28.96	0	.00	1	.32	303	99.67
37	38.54	2	2.08	57	59.37	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
284	76.34	31	8.33	57	15.32	46	79.31	3	5.17	9	15.51
26	86.66	2	6.66	2	6.66	0	.00	0	.00	0	.00
15	93.75	1	6.25	0	.00	0	.00	0	.00	0	.00
1	50.00	0	.00	1	50.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	3	100.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 15--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/E/	2538							
		E	0	.00	0	.00	1	100.00
		EE	1	100.00	0	.00	0	.00
		EA	16	94.11	0	.00	1	5.88
		E-E	3	100.00	0	.00	0	.00
		I-E	1	100.00	0	.00	0	.00
		I	0	.00	0	.00	1	100.00
		IE	0	.00	0	.00	0	.00
		EA-E	1	100.00	0	.00	0	.00
		IE-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
		EE-E	0	.00	0	.00	0	.00
		EI-E	0	.00	0	.00	0	.00
		EY	0	.00	0	.00	0	.00
		AE	0	.00	0	.00	0	.00
		OE	0	.00	0	.00	0	.00
		EO	0	.00	0	.00	0	.00
/E2/	198							
		E	1	100.00	0	.00	0	.00
		EA	4	100.00	0	.00	0	.00
		EE	2	100.00	0	.00	0	.00
		E-E	0	.00	0	.00	0	.00
		IE	1	100.00	0	.00	0	.00
		I	0	.00	0	.00	0	.00
		IE-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
/E3/	3646							
		E	212	28.30	64	8.54	473	63.15
		EA	0	.00	0	.00	0	.00
		A	2	6.25	30	93.75	0	.00
		E-E	8	53.33	0	.00	7	46.66
		AI	0	.00	0	.00	0	.00
		IE	0	.00	0	.00	0	.00
		EO	0	.00	0	.00	0	.00
		U	0	.00	0	.00	0	.00
		A-E	0	.00	0	.00	1	100.00
		AE	0	.00	0	.00	1	100.00
		AY	0	.00	0	.00	0	.00
		EA-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
/E5/	2170							
		E	0	.00	0	.00	820	100.00
		O	0	.00	0	.00	70	100.00

TABLE 15--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
18	75.00	0	.00	6	25.00	258	14.82	87	5.00	1395	80.17
153	92.16	11	6.62	2	1.20	69	84.14	3	3.65	10	12.19
167	95.42	8	4.57	0	.00	53	100.00	0	.00	0	.00
42	71.18	1	1.69	16	27.11	0	.00	0	.00	0	.00
28	65.11	1	2.32	14	32.55	0	.00	0	.00	0	.00
13	92.85	0	.00	1	7.14	15	65.21	1	4.34	7	30.43
30	90.90	2	6.06	1	3.03	0	.00	0	.00	0	.00
27	93.10	0	.00	2	6.89	0	.00	0	.00	0	.00
17	73.91	5	21.73	1	4.34	0	.00	0	.00	0	.00
11	100.00	0	.00	0	.00	5	100.00	0	.00	0	.00
9	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
6	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	5	83.33	1	16.66	0	.00
0	.00	0	.00	0	.00	2	40.00	0	.00	3	60.00
0	.00	0	.00	0	.00	5	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	50.00	1	50.00	0	.00
53	84.12	9	14.28	1	1.58	0	.00	0	.00	0	.00
44	97.77	1	2.22	0	.00	0	.00	0	.00	0	.00
32	94.11	1	2.94	1	2.94	0	.00	0	.00	0	.00
15	55.55	2	7.40	10	37.03	0	.00	0	.00	0	.00
13	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1868	74.03	275	10.89	380	15.06	1	2.27	0	.00	43	97.72
119	85.61	14	10.07	6	4.31	0	.00	0	.00	0	.00
1	1.61	58	93.54	3	4.83	0	.00	0	.00	0	.00
53	82.81	5	7.81	6	9.37	0	.00	0	.00	0	.00
3	75.00	1	25.00	0	.00	0	.00	0	.00	0	.00
4	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	840	100.00	1	16.66	0	.00	5	83.33
0	.00	0	.00	198	100.00	0	.00	0	.00	0	.00

TABLE 15--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/E5/--Continued								
		A	0	.00	0	.00	52	100.00
		U	0	.00	0	.00	8	100.00
		U-E	0	.00	0	.00	8	100.00
		I	0	.00	0	.00	1	100.00
		Y	0	.00	0	.00	2	100.00
		E-E	0	.00	0	.00	1	100.00
		OU	0	.00	0	.00	1	100.00
/I/	1482							
		I-E	6	11.53	0	.00	46	88.46
		I	2	100.00	0	.00	0	.00
		Y	0	.00	0	.00	0	.00
		IGH	0	.00	0	.00	0	.00
		IE	0	.00	0	.00	0	.00
		Y-E	0	.00	0	.00	0	.00
		EY-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
		AY	0	.00	0	.00	0	.00
		EIGH	0	.00	0	.00	0	.00
		UY	0	.00	0	.00	0	.00
		AI-E	1	100.00	0	.00	0	.00
		EY	0	.00	0	.00	0	.00
		OY	0	.00	0	.00	0	.00
/I3/	7815							
		I	192	12.50	200	13.02	1143	74.46
		Y	0	.00	0	.00	5	100.00
		I-E	0	.00	0	.00	44	100.00
		A-E	0	.00	0	.00	108	100.00
		EY	0	.00	0	.00	0	.00
		IE	0	.00	0	.00	5	100.00
		UI	0	.00	0	.00	0	.00
		AI	0	.00	0	.00	1	100.00
		EI	3	37.50	0	.00	5	62.50
		EE	0	.00	0	.00	0	.00
		E	0	.00	0	.00	0	.00
		A	0	.00	0	.00	4	100.00
		IA-E	0	.00	0	.00	3	100.00
		U	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	3	100.00
		AY	0	.00	0	.00	0	.00
		E-E	0	.00	0	.00	0	.00
		EA	0	.00	0	.00	0	.00
		EA-E	0	.00	0	.00	1	100.00

TABLE 15--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
0	.00	0	.00	116	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	23	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	15	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	7	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	2	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
313	62.22	61	12.12	129	25.64	0	.00	0	.00	0	.00
140	89.17	16	10.19	1	.63	253	64.05	41	10.37	101	25.56
2	100.00	0	.00	0	.00	97	46.41	26	12.44	86	41.14
53	67.94	24	30.76	1	1.28	10	100.00	0	.00	0	.00
8	57.14	0	.00	6	42.85	10	83.33	1	8.33	1	8.33
8	42.10	1	5.26	10	52.63	4	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	7	100.00	0	.00	0	.00
4	80.00	1	20.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	2	66.66	0	.00	1	33.33
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	3	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
1801	64.97	224	8.08	747	26.94	0	.00	0	.00	1039	100.00
101	67.78	16	10.73	32	21.47	0	.00	0	.00	1647	100.00
35	11.86	1	.33	259	87.79	0	.00	0	.00	0	.00
1	1.26	0	.00	78	98.73	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	40	100.00
0	.00	0	.00	8	100.00	0	.00	0	.00	16	100.00
10	66.66	1	6.66	4	26.66	0	.00	0	.00	1	100.00
0	.00	0	.00	14	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	3	100.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	4	100.00
2	100.00	0	.00	0	.00	0	.00	0	.00	3	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	100.00
0	.00	0	.00	1	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 15--Continued

Pho-neme	Total Occur-rence	Gra-pheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/I3/--Continued								
		IE-E	0	.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		Y-E	0	.00	0	.00	0	.00
/o/	2587							
		O	2	40.00	0	.00	3	60.00
		O-E	2	18.18	0	.00	9	81.81
		OA	8	100.00	0	.00	0	.00
		OW	4	57.14	1	14.28	2	28.57
		OU	0	.00	0	.00	0	.00
		OE	0	.00	0	.00	0	.00
		OU-E	0	.00	0	.00	0	.00
		OO	0	.00	0	.00	0	.00
		OUGH	0	.00	0	.00	0	.00
		EAU	0	.00	0	.00	0	.00
		OH	1	100.00	0	.00	0	.00
		AU	0	.00	0	.00	0	.00
		EW	0	.00	0	.00	0	.00
		OA-E	0	.00	0	.00	0	.00
		AU-E	0	.00	0	.00	0	.00
		OW-E	0	.00	0	.00	0	.00
/OI/	149							
		OI	5	62.50	0	.00	3	37.50
		OY	0	.00	0	.00	0	.00
		OI-E	0	.00	0	.00	0	.00
		OY-E	0	.00	0	.00	0	.00
/ɑU/	406							
		OU	42	60.86	8	11.59	19	27.53
		OW	2	100.00	0	.00	0	.00
		OU-E	1	100.00	0	.00	0	.00
		OUGH	0	.00	0	.00	0	.00
		OW-E	0	.00	0	.00	0	.00
/o2/	767							
		O	31	67.39	6	13.04	9	19.56
		A	15	57.69	5	19.23	6	23.07
		AU	2	22.22	1	11.11	6	66.66
		AW	2	100.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		OUGH	1	100.00	0	.00	0	.00
		AUGH	1	100.00	0	.00	0	.00
		OA	0	.00	0	.00	0	.00

TABLE 15--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	1	100.00	0	.00	0	.00	0	.00
180	74.38	26	10.74	36	14.87	413	25.35	132	8.10	1084	66.54
244	67.96	42	11.69	73	20.33	0	.00	0	.00	0	.00
80	68.96	34	29.31	2	1.72	1	50.00	0	.00	1	50.00
12	85.71	2	14.28	0	.00	46	44.66	6	5.82	51	49.51
27	93.10	1	3.44	1	3.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	10	76.92	1	7.69	2	15.38
6	60.00	1	10.00	3	30.00	0	.00	0	.00	0	.00
7	77.77	2	22.22	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	4	50.00	0	.00	4	50.00
0	.00	0	.00	0	.00	3	50.00	0	.00	3	50.00
0	.00	0	.00	0	.00	1	33.33	2	66.66	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	3	100.00
0	.00	0	.00	0	.00	3	100.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
50	70.42	6	8.45	15	21.12	12	92.30	0	.00	1	7.69
0	.00	0	.00	40	85.00	41	85.41	2	4.16	5	10.41
6	75.00	0	.00	2	25.00	0	.00	0	.00	0	.00
0	.00	0	.00	1	100.00	0	.00	0	.00	0	.00
125	81.69	21	13.72	7	4.57	5	100.00	0	.00	0	.00
40	88.88	5	11.11	0	.00	61	84.72	9	12.50	2	2.77
32	60.37	21	39.62	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	3	100.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
204	76.69	21	7.89	41	15.41	0	.00	0	.00	0	.00
102	81.60	17	13.60	6	4.80	12	85.71	2	14.28	0	.00
51	89.47	0	.00	6	10.52	53	66.25	12	15.00	15	18.75
21	95.45	0	.00	1	4.54	45	88.23	5	9.80	1	1.96
17	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
12	85.71	2	14.28	0	.00	0	.00	0	.00	0	.00
5	83.33	1	16.66	0	.00	4	80.00	1	20.00	0	.00
9	100.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 15--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/02/--Continued								
		AU-E	0	.00	0	.00	0	.00
		A-E	0	.00	0	.00	0	.00
		AO	0	.00	0	.00	0	.00
		AW-E	0	.00	0	.00	0	.00
/03/	1662							
		O	158	73.83	30	14.01	26	12.14
		A	0	.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		OW	0	.00	0	.00	0	.00
/05/	127							
		O	10	90.90	1	9.09	0	.00
		O-E	0	.00	0	.00	0	.00
/06/	453							
		OO	0	.00	0	.00	0	.00
		U	0	.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	0	.00
		OU	0	.00	0	.00	0	.00
		EW	0	.00	0	.00	0	.00
		UE	0	.00	0	.00	0	.00
		OO-E	1	100.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		UI	0	.00	0	.00	0	.00
		EU	0	.00	0	.00	0	.00
		OE	0	.00	0	.00	0	.00
		UI-E	0	.00	0	.00	0	.00
		OU-E	0	.00	0	.00	0	.00
		OUGH	0	.00	0	.00	0	.00
		UO	0	.00	0	.00	0	.00
/07/	368							
		U	0	.00	0	.00	0	.00
		OO	0	.00	0	.00	0	.00
		OU	0	.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	0	.00
		EU	0	.00	0	.00	0	.00
/ə/	6013							
		O	0	.00	0	.00	321	100.00
		A	0	.00	0	.00	6	100.00

TABLE 15--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
9	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	2	100.00	0	.00	0	.00
1056	78.62	181	13.47	106	7.89	0	.00	0	.00	0	.00
71	88.75	6	7.50	3	3.75	0	.00	0	.00	0	.00
20	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
4	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
78	69.64	13	11.60	21	18.75	0	.00	0	.00	0	.00
3	75.00	1	25.00	0	.00	0	.00	0	.00	0	.00
123	83.10	22	14.86	3	2.02	21	84.00	2	8.00	2	8.00
11	100.00	0	.00	0	.00	67	81.70	2	2.43	13	15.85
16	100.00	0	.00	0	.00	16	76.19	1	4.76	4	19.04
33	97.05	0	.00	1	2.94	0	.00	0	.00	0	.00
9	100.00	0	.00	0	.00	13	65.00	2	10.00	5	25.00
2	100.00	0	.00	0	.00	19	95.00	1	5.00	0	.00
0	.00	0	.00	0	.00	16	100.00	0	.00	0	.00
11	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
12	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
6	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	2	66.66	0	.00	1	33.33
0	.00	0	.00	0	.00	1	25.00	3	75.00	0	.00
4	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
3	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	50.00	0	.00	1	50.00
0	.00	0	.00	0	.00	2	100.00	0	.00	0	.00
46	28.04	5	3.04	113	68.90	0	.00	0	.00	36	100.00
75	65.78	23	20.17	16	14.03	0	.00	0	.00	0	.00
15	71.42	0	.00	6	28.57	0	.00	0	.00	4	100.00
8	72.72	3	27.27	0	.00	0	.00	0	.00	6	100.00
10	90.90	0	.00	1	9.09	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	1176	100.00	0	.00	0	.00	114	100.00
0	.00	0	.00	13	100.00	1	.07	0	.00	1418	99.92

TABLE 15--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/ə/--Continued								
		I	0	.00	0	.00	6	100.00
		E	0	.00	0	.00	228	100.00
		OU	0	.00	0	.00	191	100.00
		U	0	.00	0	.00	90	100.00
		E-E	0	.00	0	.00	37	100.00
		Y	0	.00	0	.00	2	100.00
		IE	0	.00	0	.00	13	100.00
		O-E	0	.00	0	.00	1	100.00
		EO	0	.00	0	.00	5	100.00
		AI	0	.00	0	.00	3	100.00
		EOU	0	.00	0	.00	0	.00
		I-E	0	.00	0	.00	2	100.00
		IE-E	0	.00	0	.00	1	100.00
		U-E	0	.00	0	.00	1	100.00
		A-E	0	.00	0	.00	1	100.00
		EA	0	.00	0	.00	2	100.00
		EI	0	.00	0	.00	1	100.00
		IA	0	.00	0	.00	0	.00
		OI-E	0	.00	0	.00	0	.00
		AU	0	.00	0	.00	0	.00
/U/	1188							
		U	7	77.77	0	.00	2	22.22
		U-E	8	25.80	0	.00	23	74.19
		EW	0	.00	0	.00	0	.00
		EU	0	.00	0	.00	0	.00
		UE	0	.00	0	.00	1	100.00
		UI	0	.00	0	.00	0	.00
		IEW	0	.00	0	.00	0	.00
		EAU	0	.00	0	.00	0	.00
		IEU	0	.00	0	.00	0	.00
		EU-E	0	.00	0	.00	0	.00
		EW-E	0	.00	0	.00	0	.00
/U2/	787							
		E	9	90.00	0	.00	1	10.00
		U	7	87.50	0	.00	1	12.50
		I	2	100.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		E-E	1	50.00	0	.00	1	50.00
		EA	15	100.00	0	.00	0	.00
		OU	0	.00	0	.00	0	.00
		U-E	1	100.00	0	.00	0	.00
		EU	0	.00	0	.00	0	.00

TABLE 15--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
0	.00	0	.00	9	100.00	0	.00	0	.00	1332	100.00
1	.23	0	.00	419	99.76	0	.00	0	.00	115	100.00
0	.00	0	.00	144	100.00	0	.00	0	.00	1	100.00
0	.00	0	.00	165	100.00	0	.00	0	.00	42	100.00
0	.00	0	.00	64	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	1	100.00	0	.00	0	.00	20	100.00
0	.00	0	.00	7	100.00	0	.00	0	.00	2	100.00
0	.00	0	.00	20	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	5	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	2	100.00	0	.00	0	.00	4	100.00
0	.00	0	.00	8	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	4	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	3	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	3	100.00
0	.00	0	.00	2	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	1	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	2	100.00
0	.00	0	.00	2	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	100.00
22	62.85	0	.00	13	37.14	265	34.31	55	7.14	450	58.44
101	44.88	0	.00	124	55.11	0	.00	0	.00	0	.00
7	100.00	0	.00	0	.00	24	77.41	1	3.22	6	19.35
1	50.00	0	.00	1	50.00	13	50.00	4	15,38	9	34.61
0	.00	0	.00	0	.00	11	42.30	0	.00	15	57.69
6	85.71	1	14.28	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	4	66.66	1	16.66	1	16.66
0	.00	0	.00	0	.00	4	80.00	0	.00	1	20.00
0	.00	0	.00	0	.00	3	75.00	0	.00	1	25.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
244	80.52	35	11.55	24	7.92	0	.00	0	.00	0	.00
171	87.69	10	5.12	14	7.17	0	.00	0	.00	0	.00
83	81.37	18	17.64	1	.98	0	.00	0	.00	0	.00
31	58.49	22	41.50	0	.00	0	.00	0	.00	0	.00
36	92.30	0	.00	3	7.69	0	.00	0	.00	0	.00
14	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
20	95.23	1	4.76	0	.00	0	.00	0	.00	0	.00
8	88.88	0	.00	1	11.11	0	.00	0	.00	0	.00
6	100.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 15--Continued

Pho-neme	Total Occur-rence	Gra-pheme	Initial					
			Primary		Secondary		unaccented	
			f	%	f	%	f	%
/U2/--Continued								
		EA-E	0	.00	0	.00	0	.00
		Y	0	.00	0	.00	0	.00
		I-E	0	.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		OU-E	0	.00	0	.00	0	.00
/U3/	1410							
		U	58	21.48	49	18.14	163	60.37
		O	5	100.00	0	.00	0	.00
		OU	0	.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	0	.00
		OO	0	.00	0	.00	0	.00

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

TABLE 15--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
822	87.44	67	7.12	51	5.42	0	.00	0	.00	2	100.00
95	88.78	10	9.34	2	1.86	0	.00	0	.00	0	.00
29	96.66	0	.00	1	3.33	0	.00	0	.00	0	.00
24	92.30	1	3.84	1	3.84	0	.00	0	.00	0	.00
22	95.65	0	.00	1	4.34	0	.00	0	.00	0	.00
7	100.00	0	.00	0	.00	0	.00	0	.00	0	.00

/O6/: This phoneme is used only in medial and final positions, with the exception of the word ooze, and occurs predominantly in primary accented syllables.

/O7/: Observe that all occurrences of this phoneme in final position are in unaccented syllables.

/ə/: By definition, all occurrences of this phoneme will appear in unaccented syllables. The data bear this out, with but two exceptions. Note < E > appearing as its graphemic option in medial position in a primary accented syllable. This is found in the word complimentary [classified as /E4/ in the 62 phoneme classification and reclassified as /ə / here] Note, as well, the single occurrence of < A > in final position in a primary accented syllable. This is the article a.

/U/: Note the infrequent use of this phoneme in secondary stressed syllables, with the exception of < U > which occurs 55 times in final position in syllables having secondary stress, and a few infrequent occurrences of < EW >, < EU >, < UI >, and < IEW >.

/U2/: Observe that this phoneme predominantly occurs in primary accented syllables, with the exception of a few occurrences in medial position in secondary accented syllables.

30 consonant classification. Consonant phonemes should be little affected by the stress patterns of American-English since syllabic stress alters primarily the quality of vowel phonemes. It would be of value to determine, however, whether the restrictions placed upon vowel phoneme-grapheme correspondences by syllabic stress also affects consonant phoneme-grapheme correspondences occurring in the same syllable. Table 16 provides the information with which to examine this question.[2]

---

[2] Table 5, pages 61-65, provides illustrations of the consonant phoneme-grapheme correspondences discussed in the text to follow. The reader may find it helpful to refer to this table occasionally as he interprets these findings.

/CH/: Observe that < TI > occurs only in unaccented syllables, as in the second syllable of nation. Otherwise, with the exception of < T >, nearly all occurrences of this phoneme are in primary accented syllables.

/F/: Note that this phoneme's two most frequently used options, < F > and < PH >, occur often in unaccented syllables.

/J/: Significant stress influences are particularly noticeable for this phoneme's less frequently used graphemic options: < D > occurs in initial position only in unaccented syllables; < DJ > occurs only in primary accented syllables; and < GI > occurs initially only in unaccented syllables and finally only in primary accented syllables.

/K/: This phoneme is used mainly in primary accented syllables, although < C > also occurs numerous times in unaccented syllables.

/NG/: Note that the option < N > seldom occurs in unaccented syllables, while < NG > occurs frequently in unaccented syllables in a final position.

/Nl/: This syllabic consonant occurs only in unaccented syllables.

/S/: This phoneme is distributed throughout all degrees of syllabic stress, although its infrequently used options in final position are confined mostly to primary accented syllables.

/SH/: When occurring in initial position, this phoneme is mainly restricted to unaccented syllables. Note, however, that < SH > and < CH > also occur numerous times in primary accented syllables. When occurring in final position, this phoneme is used mainly in primary accented syllables, with the exception of 90 occurrences of < SH > in unaccented syllables. The frequently used option < SH > is distributed somewhat randomly through all levels of stress.

TABLE 16

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN STRESSED AND UNSTRESSED SYLLABLES; 30 CONSONANT CLASSIFICATION

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/B/	2302	B	964	52.19	155	8.39	728	39.42
		BB	3	60.00	2	40.00	0	.00
		PB	0	.00	1	100.00	0	.00
/CH/	564	CH	145	81.92	6	3.38	26	14.70
		T	0	.00	0	.00	127	100.00
		TCH	0	.00	0	.00	0	.00
		TI	0	.00	0	.00	13	100.00
		C	2	100.00	0	.00	0	.00
/D/	3691	D	795	37.91	161	7.67	1141	54.42
		DD	7	100.00	0	.00	0	.00
		LD	0	.00	2	100.00	0	.00
/F/	2019	F	889	64.37	89	6.44	403	29.19
		PH	59	32.41	14	7.69	109	59.90
		FF	59	95.16	2	3.22	1	1.62
		LF	0	.00	2	40.00	3	60.00
		GH	0	.00	0	.00	0	.00
		FT	0	.00	0	.00	0	.00
/G/	1338	G	538	61.76	47	5.39	286	32.85
		GG	5	100.00	0	.00	0	.00
		X	0	.00	0	.00	0	.00
		GUE	0	.00	0	.00	0	.00
		GU	15	78.95	4	21.05	0	.00
		GH	7	7.77	0	.00	2	22.23
		TG	0	.00	0	.00	1	100.00
/H/	778	H	573	75.19	87	11.41	102	13.40
		WH	12	75.00	3	18.75	1	6.25
/HW/	89	WH	63	70.78	12	13.48	14	15.74

TABLE 16--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
1	100.00	0	.00	0	.00	238	60.86	46	11.76	107	27.38
0	.00	0	.00	0	.00	56	96.55	2	3.45	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	124	92.53	4	2.98	6	4.49
0	.00	0	.00	0	.00	35	72.91	10	20.83	3	6.26
0	.00	0	.00	0	.00	59	96.72	2	3.28	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
16	55.17	1	3.44	12	41.39	892	60.06	171	11.51	422	28.43
1	100.00	0	.00	0	.00	65	98.48	1	1.52	0	.00
0	.00	0	.00	0	.00	4	100.00	0	.00	0	.00
37	86.04	5	11.62	1	2.34	138	88.46	16	10.25	2	1.29
4	44.44	1	11.11	4	44.45	32	62.74	4	7.84	15	29.42
0	.00	0	.00	0	.00	90	78.26	16	13.91	9	7.83
0	.00	0	.00	0	.00	4	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	8	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	3	100.00	0	.00	0	.00
1	100.00	0	.00	0	.00	228	74.50	33	10.78	45	14.72
0	.00	0	.00	0	.00	59	95.16	3	4.84	0	.00
0	.00	0	.00	0	.00	0	.00	2	4.76	40	95.24
0	.00	0	.00	0	.00	10	47.61	0	.00	11	52.39
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 16--Continued

Phoneme	Total Occurrence	Grapheme	Initial Primary f	Initial Primary %	Initial Secondary f	Initial Secondary %	Initial Unaccented f	Initial Unaccented %
/H9/	41							
		H	16	76.19	2	9.52	3	14.29
/J/	982							
		G	104	30.58	15	4.41	221	65.01
		J	178	83.17	8	3.73	28	13.10
		DG	0	.00	0	.00	0	.00
		D	0	.00	0	.00	7	100.00
		DJ	13	100.00	0	.00	0	.00
		GI	0	.00	0	.00	11	100.00
		GG	0	.00	0	.00	0	.00
		DI	1	50.00	0	.00	1	50.00
/K/	4712							
		C	1204	52.64	169	7.38	914	39.98
		K	70	43.20	10	6.17	82	50.63
		CK	0	.00	0	.00	0	.00
		CH	43	44.79	5	5.20	48	50.01
		X	0	.00	0	.00	0	.00
		CC	41	89.13	4	8.69	1	2.18
		QU	5	62.50	0	.00	3	37.50
		Q	0	.00	0	.00	2	100.00
		LK	0	.00	0	.00	0	.00
		CQ	0	.00	0	.00	0	.00
		KH	2	66.67	0	.00	1	33.33
		SC	0	.00	3	100.00	0	.00
		CCH	0	.00	0	.00	0	.00
/KS/	271							
		X	0	.00	0	.00	0	.00
		CS	0	,00	0	.00	0	.00
/KW/	196							
		QU	107	64.84	6	3.63	52	31.53
		CQU	5	100.00	0	.00	0	.00
/L/	5389							
		L	875	49.46	118	6.67	776	43.87
		LL	82	78.84	9	8.65	13	12.51
		SL	0	.00	0	.00	3	100.00
		LN	0	.00	0	.00	0	.00
/L1/	651							
		L	0	.00	0	.00	0	.00
		EL	0	.00	0	.00	0	.00

TABLE 16--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
8	72.72	2	18.18	1	9.10	0	.00	0	.00	9	100.00
0	.00	0	.00	0	.00	169	55.04	16	5.21	122	39.75
0	.00	0	.00	0	.00	3	75.00	1	25.00	0	.00
0	.00	0	.00	0	.00	44	86.27	1	1.96	6	11.77
0	.00	0	.00	0	.00	20	80.00	4	16.00	1	4.00
0	.00	0	.00	0	.00	3	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	3	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	50.00	1	50.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
213	74.47	4	1.39	69	24.14	414	47.09	57	6.48	408	46.43
40	85.11	7	14.89	0	.00	320	81.63	58	14.79	14	3.58
1	100.00	0	.00	0	.00	228	78.89	35	12.11	26	9.00
9	69.23	1	7.69	3	23.08	13	39.39	9	27.27	11	33.34
0	.00	0	.00	0	.00	60	75.00	15	18.75	5	6.25
0	.00	0	.00	0	.00	23	76.66	7	23.33	0	.00
0	.00	0	.00	0	.00	19	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	16	88.89	2	11.11	0	.00
0	.00	0	.00	0	.00	11	78.57	3	21.43	0	.00
0	.00	0	.00	0	.00	0	.00	3	100.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
6	75.00	0	.00	2	25.00	63	26.58	22	9.28	152	64.14
0	.00	0	.00	0	.00	0	.00	0	.00	26	100.00
26	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
885	79.87	63	5.68	160	14.45	919	45.56	129	6.39	969	48.05
0	.00	1	100.00	0	.00	323	84.11	35	9.11	26	6.78
0	.00	0	.00	0	.00	2	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	1	100.00	0	.00	0	.00	619	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	19	100.00

TABLE 16--Continued

Pho- neme	Total Occur- rence	Gra- pheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/Ll/--Continued								
		IL	0	.00	0	.00	0	.00
		AL	0	.00	0	.00	0	.00
		OL	0	.00	0	.00	0	.00
/M/	3501							
		M	956	51.01	160	8.53	758	40.46
		MM	53	86.88	5	8.19	3	4.93
		MB	0	.00	0	.00	0	.00
		LM	0	.00	0	.00	2	100.00
		GM	0	.00	1	20.00	4	80.00
		MN	0	.00	0	.00	0	.00
/Ml/	97							
		M	0	.00	0	.00	0	.00
/N/	7656							
		N	539	44.07	64	5.23	620	50.70
		NN	25	83.33	3	10.00	2	6.67
		KN	36	87.80	2	4.87	3	7.33
		GN	5	100.00	0	.00	0	.00
		PN	0	.00	1	33.33	2	66.67
		MN	0	.00	0	.00	1	100.00
/NG/	615							
		NG	0	.00	0	.00	0	.00
		N	0	.00	0	.00	0	.00
		NGUE	0	.00	0	.00	0	.00
/Nl/	128							
		EN	0	.00	0	.00	0	.00
		ON	0	.00	0	.00	0	.00
		IN	0	.00	0	.00	0	.00
/P/	3449							
		P	1350	58.64	197	8.55	755	32.81
		PP	62	93.94	4	6.06	0	.00
/R/	9390							
		R	661	40.10	135	8.19	852	51.71
		RR	50	100.00	0	.00	0	.00
		WR	43	89.58	5	10.42	0	.00
		RH	12	75.00	1	6.25	3	18.75

TABLE 16--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
0	.00	0	.00	0	.00	0	.00	0	.00	7	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	4	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	100.00
111	93.28	4	3.36	4	3.36	727	55.53	133	10.16	449	34.31
0	.00	0	.00	0	.00	67	84.81	11	13.92	1	1.27
0	.00	0	.00	0	.00	24	88.88	2	7.40	1	3.72
2	100.00	0	.00	0	.00	13	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	33.33	0	.00	2	66.67
0	.00	0	.00	0	.00	4	57.14	0	.00	3	42.86
0	.00	0	.00	0	.00	6	6.18	0	.00	91	93.82
499	36.47	52	3.80	817	59.73	1785	36.72	436	8.96	2640	54.32
0	.00	0	.00	0	.00	90	92.78	6	6.18	1	1.04
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	22	81.48	0	.00	5	18.52
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
5	100.00	0	.00	0	.00	92	25.77	11	3.08	254	71.15
116	89.92	2	1.55	11	8.53	103	84.42	12	9.83	7	5.75
0	.00	0	.00	0	.00	2	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	84	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	41	100.00
0	.00	0	.00	0	.00	0	.00	0	.00	3	100.00
224	80.86	15	5.41	38	13.73	534	74.47	84	11.71	99	13.82
1	100.00	0	.00	0	.00	65	75.58	20	23.25	1	1.17
2071	65.22	244	7.68	860	27.10	1612	37.52	294	6.84	2390	55.64
0	.00	0	.00	0	.00	133	84.71	23	14.64	1	.65
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 16--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/S/	6326							
		S	1713	65.03	236	8.95	685	26.02
		C	244	45.43	24	4.46	269	50.11
		SS	53	75.71	12	17.14	5	7.14
		SC	32	82.05	3	7.69	4	10.26
		X	4	5.19	0	.00	73	94.81
		ST	0	.00	0	.00	0	.00
		PS	8	42.10	5	26.31	6	31.59
		Z	0	.00	0	.00	2	100.00
		SW	1	25.00	1	25.00	2	50.00
		SCH	1	50.00	0	.00	1	50.00
/SH/	1537							
		TI	0	.00	0	.00	729	100.00
		SH	152	76.00	26	13.00	22	11.00
		CI	0	.00	0	.00	39	100.00
		SSI	0	.00	0	.00	0	.00
		SI	0	.00	0	.00	38	100.00
		C	0	.00	0	.00	22	100.00
		CH	15	51.72	3	10.34	11	37.94
		T	0	.00	0	.00	24	100.00
		S	8	42.10	0	.00	11	57.90
		SS	5	100.00	0	.00	0	.00
		SC	1	25.00	0	.00	3	75.00
		SCI	0	.00	0	.00	4	100.00
		X	0	.00	0	.00	3	100.00
		CE	0	.00	0	.00	2	100.00
		SCH	1	50.00	0	.00	1	50.00
/T/	7793							
		T	1329	35.70	183	4.91	2210	59.39
		TT	27	96.42	0	.00	1	3.58
		ED	0	.00	0	.00	0	.00
		BT	0	.00	0	.00	0	.00
		TW	3	100.00	0	.00	0	.00
		PT	1	100.00	0	.00	0	.00
		CT	0	.00	0	.00	0	.00
		CHT	0	.00	1	100.00	0	.00
		TH	1	100.00	0	.00	0	.00
/T1/	411							
		TH	141	62.38	15	6.63	70	30.99
/T2/	149							
		TH	21	34.42	5	8.19	35	57.39

TABLE 16--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
220	58.98	17	4.55	136	36.47	574	36.05	134	8.41	884	55.54
5	50.00	1	10.00	4	40.00	239	45.96	20	3.84	261	50.20
2	100.00	0	.00	0	.00	217	58.64	18	4.86	135	36.50
0	.00	0	.00	0	.00	33	86.84	4	10.52	1	2.64
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	33	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	4	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	89	97.80	2	2.20	0	.00
0	.00	0	.00	4	100.00	85	43.81	19	9.79	90	46.40
0	.00	0	.00	0	.00	41	97.62	1	2.38	0	.00
0	.00	0	.00	0	.00	51	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	16	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	5	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	5	83.33	1	16.67	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	4	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	2	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
366	69.58	62	11.78	98	18.64	1465	44.66	201	6.12	1614	49.22
0	.00	0	.00	0	.00	182	96.80	4	2.12	2	1.08
0	.00	0	.00	0	.00	18	64.28	1	3.57	9	32.15
0	.00	0	.00	0	.00	11	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	2	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	141	76.21	24	12.97	20	10.82
2	50.00	2	50.00	0	.00	71	84.52	3	3.57	10	11.91

TABLE 16--Continued

Pho-neme	Total Occur-rence	Gra-pheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/V/	1492							
		V	483	57.63	36	4.29	319	38.08
		LV	0	.00	0	.00	0	.00
		F	0	.00	0	.00	0	.00
/W/	626							
		W	344	69.63	95	19.23	55	11.14
		U	0	.00	0	.00	22	100.00
		OU	0	.00	0	.00	1	100.00
/Y/	120							
		I	1	1.56	0	.00	63	98.44
		Y	39	73.58	5	9.43	9	16.99
		E	0	.00	0	.00	1	100.00
/Z/	995							
		S	89	52.97	1	.59	78	46.44
		Z	49	58.33	2	2.38	33	39.29
		ES	0	.00	0	.00	0	.00
		X	36	85.71	4	9.52	2	4.77
		ZZ	1	100.00	0	.00	0	.00
		SS	11	100.00	0	.00	0	.00
		SC	3	100.00	0	.00	0	.00
		CZ	1	100.00	0	.00	0	.00
/ZH/	102							
		SI	0	.00	0	.00	35	100.00
		S	1	5.26	0	.00	18	94.74
		G	2	66.67	0	.00	1	33.33
		Z	0	.00	0	.00	1	100.00
		TI	0	.00	0	.00	1	100.00

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

TABLE 16--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
15	93.75	0	.00	1	6.25	405	64.18	35	5.54	191	30.28
2	100.00	0	.00	0	.00	2	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	3	100.00	0	.00	0	.00
83	98.81	1	1.19	0	.00	0	.00	0	.00	0	.00
12	48.00	0	.00	13	52.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
5	5.20	0	.00	91	94.80	281	74.73	37	9.84	58	15.43
0	.00	0	.00	0	.00	50	34.48	8	5.52	87	60.00
0	.00	0	.00	0	.00	35	79.54	3	6.81	6	13.65
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	22	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	2	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	14	93.33	1	6.67	0	.00
0	.00	0	.00	0	.00	15	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	10	83.33	0	.00	2	16.67
0	.00	0	.00	0	.00	1	100.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00

/Y/: When used in initial position, < I > occurs 98.44% of the time in unaccented syllables, while < Y > occurs 73.48% of the time in primary accented syllables.

/Z/: This phoneme occurs mainly in primary accented syllables, although < S > and < Z > have numerous occurrences in unaccented syllables.

/ZH/: Note that this phoneme occurs in initial position mainly in unaccented syllables, and in final position mainly in primary accented syllables.

It is evident from an anlysis of the effects of syllabic stress on the selection of phoneme-grapheme correspondences that much more can be stated about their graphemic options. Obviously, the great bulk of such effects upon the selection of graphemic options is related primarily to the vowel phonemes. Yet, certain noticeable tendencies of syllable stress toward restricting some consonant graphemic options to particular syllables can be observed.

When a summary of these effects is made for the vowel phonemes organized in a 22 vowel classification, the influences of stress are especially noticeable for /A2/, /E2/, /I3/, /O6/, /O7/, and /ə/, while particular graphemic options for the phonemes /A/, /E/, /I/, /O/, and /U/ are seen to be restricted to certain positions within syllables when syllabic stress is considered to be a factor.

Among the consonant phonemes, /CH/, /N1/, /SH/, and /ZH/ are similarly restricted to occurrences in syllables having particular amounts of stress. /N1/, being a syllabic consonant, occurs only in unaccented syllables, while /CH/, /SH/, and /ZH/ are commonly associated with /ə/ in forming the unaccented syllable heard in such words as question, social, and abrasion. In addition, certain graphemic options for /CH/, /F/, /J/, /NG/, and /Y/ tend to be restricted to particular positions in syllables because of the influence of syllabic stress.

As with the interpretation of the effects of position on phoneme-grapheme correspondences, the purpose of examining the influences of stress on these relationships is not to provide a detailed analysis of each phoneme-grapheme correspondence in the corpus. Rather, the intention has been to determine whether or not syllabic stress has possible relevance for the central undertaking of this research: a determination of which factors in the phonological structure underlying the orthography are most productive in determining how nearly the orthography approximates the alphabetic principle.

It may be concluded, from the interpretations made thus far, that as this analysis proceeds further into the influences of the phonological structure upon the selection of phoneme-grapheme correspondences, decreasingly productive inferences can be made. It becomes apparent that positional factors singly provide much more relevant information concerning these influences than do additional considerations of the effects of syllabic stress.

The Distribution of Phoneme-Grapheme Correspondences in American-English

This analysis of the phonological factors affecting the orthography has been directed toward an examination of phoneme-grapheme correspondences irrespective of possible relationships among the graphemic options for a particular phoneme. Each graphemic option has been, in effect, considered in terms of its position in syllables, without considering the distribution of all the graphemic options for a particular phoneme in any one position.

To illustrate, the consonant phoneme /CH/ has five graphemic representations. It has been seen in Table 14, page 105, that < CH >, the most frequently used option for /CH/, is used 56.54% of the time in initial position, .63% of the time in medial position, and 42.83% of the time in final position in syllables. Such information is useful, but, more importantly perhaps, it would be helpful to know how many times < CH > is used in each position in relation to the other options for this phoneme: < T >, < TCH >, < TI >, and < C >. It is this kind of relationship which enables the individual to choose the correct graphemic option for a particular phoneme in spelling, since it is entirely possible that all five of these options may occur in the American-English orthography in the same position. If it can be determined that < CH >, for example, is the only graphemic option of the five that will occur in initial position in syllables, the speller then can eliminate the other options when this phoneme occurs in this position. In effect, it would be highly relevant to determine how all the graphemic options for a given phoneme are distributed among each position in syllables, as well as knowing how each graphemic option is distributed among all the positions in syllables.

Accordingly, the frequency and percentage of occurrence of all the graphemic options for a given phoneme in terms of their distribution in particular positions in syllables was tabulated. Although the data to follow provide this information for all phoneme-grapheme correspondence in the corpus, the analysis of the distribution of graphemic options in each position in syllables will be concerned only with those phonemes that fall below the 80% criterion in order to determine whether a particular option exceeds this criterion.

Table 17 and those to follow indicate the distribution of graphemic options for each phoneme for each position in syllables. These tables include:

1. The phoneme classified.
2. The number of occurrences of the phoneme in the corpus.
3. The different graphemic options used which represent the phoneme.
4. The frequency and percentage tabulations of the distribution of each option in initial position in syllables.
5. The frequency and percentage tabulations of the distribution of each option in medial position in syllables.
6. The frequency and percentage tabulations of the distribution of each option in final position in syllables.

The distribution of phoneme-grapheme correspondences by position in syllables: 22 vowel classification. /A/: As seen in Table 17, this phoneme has 16 different graphemic representations, with < A >, < A-E >, and < AI > representing /A/ 88.96% of the time (as seen previously in Table 11, page 88. < A > represents /A/ 80.98% of the time in final position, while < A-E > represents it 87.05% of the time in initial position.

TABLE 17

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF THE DISTRIBUTION OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN SYLLABLES: 22 VOWEL CLASSIFICATION

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/A/	2248		139		1047		1062	
		A	11	7.91	131	12.51	860	80.98
		A-E	121	87.05	669	63.90	0	.00
		AI	5	3.60	187	17.86	16	1.51
		AY	0	.00	3	.29	128	12.05
		AI-E	1	.72	17	1.62	0	.00
		EIGH	1	.72	7	.67	10	.94
		E	0	.00	0	.00	16	1.51
		EA	0	.00	13	1.24	1	.09
		EI	0	.00	11	1.05	3	.28
		EY	0	.00	0	.00	14	1.32
		ET	0	.00	0	.00	9	.85
		E-E	0	.00	2	.19	4	.38
		AIGH	0	.00	4	.38	0	.00
		EI-E	0	.00	2	.19	0	.00
		AU-E	0	.00	1	.10	0	.00
		AY-E	0	.00	0	.00	1	.09
/A2/	220		18		202		0	
		A	1	5.56	63	31.19	0	.00
		A-E	0	.00	50	24.75	0	.00
		AI	9	50.00	37	18.32	0	.00
		E-E	1	5.56	33	16.34	0	.00
		EA	0	.00	13	6.44	0	.00
		EI	3	16.67	2	.99	0	.00
		E	0	.00	4	1.98	0	.00
		AI-E	3	16.67	0	.00	0	.00
		EY	1	5.56	0	.00	0	.00
/A3/	4340		1060		2976		304	
		A	1009	95.19	2879	96.74	304	100.00
		A-E	51	4.81	96	3.23	0	.00
		AI	0	.00	1	.03	0	.00
/A5/	580		95		424		61	
		A	88	92.63	373	87.74	58	95.08
		A-E	1	1.05	30	7.08	0	.00
		EA	2	2.11	16	3.77	0	.00
		E	3	3.16	2	.47	0	.00
		AH	0	.00	1	.24	3	4.92
		AU	1	1.05	3	.71	0	.00

TABLE 17--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial		Medial		Final	
			f	%	f	%	f	%
/E/	2538		25		592		1921	
		E	1	4.00	24	4.05	1740	90.58
		EE	1	4.00	166	28.04	82	4.27
		EA	17	68.00	175	29.56	53	2.76
		E-E	3	12.00	59	9.97	0	.00
		I-E	1	4.00	43	7.26	0	.00
		I	1	4.00	14	2.36	23	1.20
		IE	0	.00	33	5.57	0	.00
		EA-E	1	4.00	29	4.90	0	.00
		IE-E	0	.00	23	3.89	0	.00
		EI	0	.00	11	1.86	5	.26
		EE-E	0	.00	9	1.52	0	.00
		EI-E	0	.00	6	1.01	0	.00
		EY	0	.00	0	.00	6	.31
		AE	0	.00	0	.00	5	.26
		OE	0	.00	0	.00	5	.26
		EO	0	.00	0	.00	2	.10
/E2/	198		8		190		0	
		E	1	12.50	63	33.16	0	.00
		EA	4	50.00	45	23.68	0	.00
		EE	2	25.00	34	17.89	0	.00
		E-E	0	.00	27	14.21	0	.00
		IE	1	12.50	13	6.84	0	.00
		I	0	.00	3	1.58	0	.00
		IE-E	0	.00	3	1.58	0	.00
		EI	0	.00	2	1.05	0	.00
/E3/	3646		798		2804		44	
		E	749	93.86	2523	89.98	44	100.00
		EA	0	.00	139	4.96	0	.00
		A	32	4.01	62	2.21	0	.00
		E-E	15	1.88	64	2.28	0	.00
		AI	0	.00	4	.14	0	.00
		IE	0	.00	4	.14	0	.00
		EO	0	.00	3	.11	0	.00
		U	0	.00	2	.07	0	.00
		A-E	1	.13	0	.00	0	.00
		AE	1	.13	0	.00	0	.00
		AY	0	.00	1	.04	0	.00
		EA-E	0	.00	1	.04	0	.00
		EI	0	.00	1	.04	0	.00
/E5/	2170		963		1201		6	
		E	820	85.15	840	69.94	6	100.00
		O	70	7.27	198	16.49	0	.00

TABLE 17--Continued

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/E5/--Continued								
		A	52	5.40	116	9.66	0	.00
		U	8	.83	23	1.92	0	.00
		U-E	8	.83	15	1.25	0	.00
		I	1	.10	7	.58	0	.00
		Y	2	.21	2	.17	0	.00
		E-E	1	.10	0	.00	0	.00
		OU	1	.10	0	.00	0	.00
/I/	1482		55		781		646	
		I-E	52	94.55	503	64.60	0	.00
		I	2	3.64	157	20.10	395	61.15
		Y	0	.00	2	.26	209	32.35
		IGH	0	.00	78	9.99	10	1.55
		IE	0	.00	14	1.79	12	1.86
		Y-E	0	.00	19	2.43	4	.62
		EY-E	0	.00	0	.00	7	1.08
		EI	0	.00	5	.64	1	.15
		AY	0	.00	0	.00	3	.46
		EIGH	0	.00	3	.38	0	.00
		UY	0	.00	0	.00	3	.46
		AI-E	1	1.82	0	.00	0	.00
		EY	0	.00	0	.00	1	.15
		OY	0	.00	0	.00	1	.15
/I3/	7815		1717		3346		2752	
		I	1535	89.40	2772	82.85	1039	37.75
		Y	5	.29	149	4.45	1647	59.85
		I-E	44	2.56	292	8.82	0	.00
		A-E	108	6.29	79	2.36	0	.00
		EY	0	.00	0	.00	40	1.45
		IE	5	.29	8	.24	16	.58
		UI	0	.00	15	.45	1	.04
		AI	1	.06	14	.42	0	.00
		EI	8	.46	3	.09	0	.00
		EE	0	.00	2	.06	4	.15
		E	0	.00	2	.06	3	.11
		A	4	.23	0	.00	0	.00
		IA-E	3	.17	0	.00	0	.00
		U	0	.00	3	.09	0	.00
		U-E	3	.17	0	.00	0	.00
		AY	0	.00	0	.00	1	.04
		E-E	0	.00	1	.03	0	.00
		EA	0	.00	0	.00	1	.04
		EA-E	1	.06	0	.00	0	.00
		IE-E	0	.00	1	.03	0	.00
		O	0	.00	1	.03	0	.00
		Y-E	0	.00	1	.03	0	.00

TABLE 17--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial f	Initial %	Medial f	Medial %	Final f	Final %
/O/	2587		32		784		1771	
		O	5	15.63	242	30.87	1629	91.98
		O-E	11	34.38	359	45.79	0	.00
		OA	8	25.00	116	14.80	2	.11
		OW	7	21.88	14	1.79	103	5.82
		OU	0	.00	29	3.70	0	.00
		OE	0	.00	0	.00	13	.73
		OU-E	0	.00	10	1.28	0	.00
		OO	0	.00	9	1.15	0	.00
		OUGH	0	.00	0	.00	8	.45
		EAU	0	.00	0	.00	6	.34
		OH	1	3.13	0	.00	3	.17
		AU	0	.00	0	.00	3	.17
		EW	0	.00	0	.00	3	.17
		OA-E	0	.00	3	.38	0	.00
		AU-E	0	.00	2	.26	0	.00
		OW-E	0	.00	0	.00	1	.06
/OI/	149		8		80		61	
		OI	8	100.00	71	88.75	13	21.31
		OY	0	.00	0	.00	48	78.69
		OI-E	0	.00	8	10.00	0	.00
		OY-E	0	.00	1	1.25	0	.00
/OU/	406		72		254		80	
		OU	69	95.83	153	60.24	5	6.25
		OW	2	2.78	45	17.72	72	90.00
		OU-E	1	1.39	53	20.87	0	.00
		OUGH	0	.00	1	.39	3	3.75
		OW-E	0	.00	2	.79	0	.00
/O2/	767		85		530		152	
		O	46	54.12	266	50.19	0	.00
		A	26	30.59	125	23.58	14	9.21
		AU	9	10.59	57	10.75	80	52.63
		AW	2	2.35	22	4.15	51	33.55
		O-E	0	.00	17	3.21	0	.00
		OUGH	1	1.18	14	2.64	0	.00
		AUGH	1	1.18	6	1.13	5	3.29
		OA	0	.00	9	1.70	0	.00
		AU-E	0	.00	9	1.70	0	.00
		A-E	0	.00	3	.57	0	.00
		AO	0	.00	2	.38	0	.00
		AW-E	0	.00	0	.00	2	1.32

TABLE 17--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial		Medial		Final	
			f	%	f	%	f	%
/03/	1662		214		1447		1	
		O	214	100.00	1343	92.81	1	100.00
		A	0	.00	80	5.53	0	.00
		O-E	0	.00	20	1.38	0	.00
		OW	0	.00	4	.28	0	.00
/05/	127		11		116		0	
		O	11	100.00	112	96.55	0	.00
		O-E	0	.00	4	3.45	0	.00
/06/	453		1		257		195	
		OO	0	.00	148	57.59	25	12.82
		U	0	.00	11	4.28	82	42.05
		O	0	.00	16	6.23	21	10.77
		U-E	0	.00	34	13.23	0	.00
		OU	0	.00	9	3.50	20	10.26
		EW	0	.00	2	.78	20	10.26
		UE	0	.00	0	.00	16	8.21
		OO-E	1	100.00	11	4.28	0	.00
		O-E	0	.00	12	4.67	0	.00
		UI	0	.00	6	2.33	0	.00
		EU	0	.00	1	.39	3	1.54
		OE	0	.00	0	.00	4	2.05
		UI-E	0	.00	4	1.56	0	.00
		OU-E	0	.00	3	1.17	0	.00
		OUGH	0	.00	0	.00	2	1.03
		UO	0	.00	0	.00	2	1.03
/07/	368		0		322		46	
		U	0	.00	164	50.93	36	78.26
		OO	0	.00	114	35.40	0	.00
		OU	0	.00	21	6.52	4	8.70
		O	0	.00	11	3.42	6	13.04
		U-E	0	.00	11	3.42	0	.00
		EU	0	.00	1	.31	0	.00
/ə/	6013		911		2046		3056	
		O	321	35.24	1176	57.48	114	3.73
		A	6	.66	13	.64	1419	46.43
		I	6	.66	9	.44	1332	43.59
		E	228	25.03	420	20.53	115	3.76
		OU	191	20.97	144	7.04	1	.03
		U	90	9.88	165	8.06	42	1.37
		E-E	37	4.06	64	3.13	0	.00
		Y	2	.22	1	.05	20	.65
		IE	13	1.43	7	.34	2	.07
		O-E	1	.11	20	.98	0	.00
		EO	5	.55	5	.24	0	.00

TABLE 17--Continued

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/ə/--Continued								
		AI	3	.33	2	.10	4	.13
		EOU	0	.00	8	.39	0	.00
		I-E	2	.22	4	.20	0	.00
		IE-E	1	.11	3	.15	0	.00
		U-E	1	.11	0	.00	3	.10
		A-E	1	.11	2	.10	0	.00
		EA	2	.22	1	.05	0	.00
		EI	1	.11	0	.00	1	.03
		IA	0	.00	0	.00	2	.07
		OI-E	0	.00	2	.10	0	.00
		AU	0	.00	0	.00	1	.03
/U/	1188		41		277		870	
		U	9	21.95	35	12.64	770	88.51
		U-E	31	75.60	225	81.23	0	.00
		EW	0	.00	7	2.53	31	3.56
		EU	0	.00	2	.72	26	2.99
		UE	1	2.44	0	.00	26	2.99
		UI	0	.00	7	2.53	1	.11
		IEW	0	.00	0	.00	6	.69
		EAU	0	.00	0	.00	5	.57
		IEU	0	.00	0	.00	4	.46
		EU-E	0	.00	1	.36	0	.00
		EW-E	0	.00	0	.00	1	.11
/U2/	787		38		749		0	
		E	10	26.32	303	40.45	0	.00
		U	8	21.05	195	26.03	0	.00
		I	2	5.26	102	13.62	0	.00
		O	0	.00	53	7.08	0	.00
		E-E	2	5.26	39	5.21	0	.00
		OU	0	.00	21	2.80	0	.00
		U-E	1	2.63	9	1.20	0	.00
		EU	0	.00	6	.80	0	.00
		EA-E	0	.00	2	.27	0	.00
		Y	0	.00	2	.27	0	.00
		I-E	0	.00	1	.13	0	.00
		O-E	0	.00	1	.13	0	.00
		OU-E	0	.00	1	.13	0	.00
/U3/	1410		275		1133		2	
		U	270	98.18	940	82.97	2	100.00
		O	5	1.82	107	9.44	0	.00
		OU	0	.00	30	2.65	0	.00

TABLE 17--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial f	Initial %	Medial f	Medial %	Final f	Final %
/U3/--Continued								
		O-E	0	.00	26	2.29	0	.00
		U-E	0	.00	23	2.03	0	.00
		OO	0	.00	7	.62	0	.00

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

/A2/: No graphemic option exceeds the 80% criterion in any position in syllables.

/E/: Sixteen different graphemic options are used to represent this phoneme. Three options, < E >, < EE >, and < EA > do so 89.00% of the time. Note that < E > represents this phoneme 90.58% of the time in final position.

/E2/: No graphemic option exceeds the 80% criterion in any position in syllables.

/E5/: Nine graphemic options are used to represent /E5/. < E > and < O > do so 89.12% of the time. < E >, however, represants /E5/ 85.15% of the time in initial position and 100% of the time in final position.

/I/: This phoneme is represented by 14 different graphemic options. Three options, < I-E >, < I >, and < Y >, do so 89.05% of the time. Of the 14, < I-E > represents /I/ 94.55% of the time in initial position.

/I3/: Twenty-two different graphemic options represent this phoneme. Of them, two options, < I > and < Y >, represent /I3/ 91.44%

of the time. Observe that < I > represents this phoneme 89.40% of the time in initial position and 82.85% of the time in medial positinn in syllables.

/O/: Sixteen different graphemic options are used to represent this phoneme. < O > and < O-E > do so 86.81% of the time. < O >, however, represents /O/ in final position in syllables 91.98% of the time.

/OI/: This phoneme has four graphemic representations, with two of them, < OI > and < OY >, representing /OI/ 94.05% of the time. The graphemic option < OI >, however, represents this phoneme 100% of the time in initial position (with the exception of oyster) and 88.75% of the time in medial position.

/OU/: Five different graphemic options are used to represent /OU/. Of them, < OU > and < OW > do so 85.22% of the time. Note, however, that < OU > represents this phoneme 95.93% of the time in initial position, while < OW > represents it 90.00% of the time in final position.

/O2/: No graphemic option exceeds the 80% criterion in any position in syllables.

/O6/: Excluding the single occurrence of < OO-E > in initial position (in the word ooze), no graphemic option exceeds the 80% criterion in any position in syllables.

/O7/: No graphemic option exceeds the 80% criterion in any position in syllables.

/θ/: No graphemic option exceeds the 80% criterion in any position in syllables.

/U/: Eleven different options are used to represent /U/. Of them, < U > and < U-E > do so 90.05% of the time. However, < U > represents this phoneme 88.51% of the time in final position, and < U-E >

represents /U/ 81.23% of the time in medial position.

/U2/: No graphemic option exceeds the 80% criterion in any position in syllables.

30 consonant classification. The analysis of the distribution of consonant phoneme-grapheme correspondences in Table 18 will be confined to initial and final positions in syllables only, although the data show many graphemic options exceeding the 80% criterion in medial position. All consonant phonemes occurring in medial position use their most frequent graphemic option in this position above the 80% criterion, with the exception of /NG/ and /SH/ which are represented by their second most frequently occurring options in that position. In addition, /W/ is listed as using < W > and < U > in medial position although neither option exceeds the 80% criterion.

/CH/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

/F/: This phoneme is represented by six different graphemic options, with < F > and < PH > doing so 90.27% of the time. < F > represents this phoneme in initial position 84.72% of the time.

/J/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

/K/: Thirteen different graphemic options represent this phoneme, two of them, < C > and < K >, doing so 86.00% of the time. In initial position in syllables, < C > represents /K/ 87.72% of the time.

/NG/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

/N1/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

TABLE 18

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF THE DISTRIBUTION OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN SYLLABLES: 30 CONSONANT CLASSIFICATION

Phoneme	Total Occurrence	Grapheme	Initial f	Initial %	Medial f	Medial %	Final f	Final %
/B/	2303		1853		1		449	
		B	1847	99.67	1	100.00	391	87.08
		BB	5	.27	0	.00	58	12.92
		PB	1	.05	0	.00	0	.00
/CH/	564		319		2		243	
		CH	177	55.49	2	100.00	134	55.14
		T	127	39.81	0	.00	48	19.75
		TCH	0	.00	0	.00	61	25.10
		TI	13	4.08	0	.00	0	.00
		C	2	.63	0	.00	0	.00
/D/	3691		2106		30		1555	
		D	2097	99.57	29	96.67	1485	95.50
		DD	7	.33	1	3.33	66	4.24
		LD	2	.10	0	.00	4	.26
/F/	2019		1630		52		337	
		F	1381	84.72	43	82.69	156	46.29
		PH	182	11.17	9	17.31	51	15.13
		FF	62	3.80	0	.00	115	34.13
		LF	5	.31	0	.00	4	1.19
		GH	0	.00	0	.00	8	2.37
		FT	0	.00	0	.00	3	.89
/G/	1338		905		1		432	
		G	871	96.24	1	100.00	306	70.83
		GG	5	.55	0	.00	62	14.35
		X	0	.00	0	.00	42	9.72
		GUE	0	.00	0	.00	21	4.86
		GH	9	.99	0	.00	1	.23
		TG	1	.11	0	.00	0	.00
/H/	778		778		0		0	
		H	762	97.94	0	.00	0	.00
		WH	16	2.06	0	.00	0	.00
/HW/	89		89		0		0	
		WH	89	100.00	0	.00	0	.00
/H9/	41		21		11		9	
		H	21	100.00	11	100.00	9	100.00

TABLE 18--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial		Medial		Final	
			f	%	f	%	f	%
/J/	982		587		0		395	
		G	340	57.92	0	.00	307	77.72
		J	214	36.46	0	.00	4	1.01
		DG	0	.00	0	.00	51	12.91
		D	7	1.19	0	.00	25	6.33
		DJ	13	2.22	0	.00	3	.76
		GI	11	1.87	0	.00	3	.76
		GG	0	.00	0	.00	2	.51
		DI	2	.34	0	.00	0	.00
/K/	4712		2607		347		1758	
		C	2287	87.72	286	82.42	879	50.00
		K	162	6.21	47	13.54	392	22.30
		CK	0	.00	1	.29	289	16.44
		CH	96	3.68	13	3.75	33	1.88
		X	0	.00	0	.00	80	4.55
		CC	46	1.76	0	.00	30	1.71
		QU	8	.31	0	.00	19	1.08
		Q	2	.08	0	.00	18	1.02
		LK	0	.00	0	.00	14	.80
		CQ	0	.00	0	.00	3	.17
		KH	3	.12	0	.00	0	.00
		SC	3	.12	0	.00	0	.00
		CCH	0	.00	0	.00	1	.06
/KS/	271		0		8		263	
		X	0	.00	8	100.00	237	90.11
		CS	0	.00	0	.00	26	9.89
/KW/	196		170		26		0	
		QU	165	97.06	26	100.00	0	.00
		CQU	5	2.94	0	.00	0	.00
/L/	5389		1876		1109		2404	
		L	1769	94.30	1108	99.91	2017	83.90
		LL	104	5.54	1	.09	384	15.97
		SL	3	.16	0	.00	2	.08
		LN	0	.00	0	.00	1	.05
/Ll/	651		0		1		650	
		LE	0	.00	1	100.00	619	95.23
		EL	0	.00	0	.00	19	2.92
		IL	0	.00	0	.00	7	1.08
		AL	0	.00	0	.00	4	.62
		OL	0	.00	0	.00	1	.15

TABLE 18--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial		Medial		Final	
			f	%	f	%	f	%
/M/	3501		1942		121		1438	
		M	1874	96.50	119	98.35	1309	91.03
		MM	61	3.14	0	.00	79	5.50
		MB	0	.00	0	.00	27	1.88
		LM	2	.10	2	1.65	13	.90
		GM	5	.26	0	.00	3	.21
		MN	0	.00	0	.00	7	.49
/Ml/	97		0		0		97	
		M	0	.00	0	.00	97	100.00
/N/	7656		1303		1368		4985	
		N	1223	93.86	1368	100.00	4861	97.51
		NN	30	2.30	0	.00	97	1.95
		KN	41	3.15	0	.00	0	.00
		GN	5	.38	0	.00	27	.54
		PN	3	.23	0	.00	0	.00
		MN	1	.08	0	.00	0	.00
/NG/	615		0		134		481	
		NG	0	.00	5	3.73	357	74.22
		N	0	.00	129	96.27	122	25.36
		NGUE	0	.00	0	.00	2	.42
/Nl/	128		0		0		128	
		EN	0	.00	0	.00	84	65.63
		ON	0	.00	0	.00	3	2.34
		IN	0	.00	0	.00	41	32.03
/P/	3449		2368		278		803	
		P	2302	97.12	277	99.64	717	89.29
		PP	66	2.79	1	.36	86	10.71
/R/	9390		1762		3175		4453	
		R	1648	93.53	3175	100.00	4296	96.47
		RR	50	2.84	0	.00	157	3.53
		WR	48	2.72	0	.00	0	.00
		RH	16	.91	0	.00	0	.00
/S/	6326		3384		385		2557	
		S	2634	77.84	373	96.88	1592	62.26
		C	537	15.87	10	2.60	520	20.34
		SS	70	2.07	2	.52	370	14.47
		SC	39	1.15	0	.00	38	1.49
		X	77	2.27	0	.00	0	.00
		ST	0	.00	0	.00	33	1.29

TABLE 18--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial		Medial		Final	
			f	%	f	%	f	%
/S/--Continued								
		PS	19	.56	0	.00	0	.00
		Z	2	.06	0	.00	4	.16
		SW	4	.12	0	.00	0	.00
		SCH	2	.06	0	.00	0	.00
/SH/	1537		1120		4		413	
		TI	729	65.09	0	.00	91	22.03
		SH	200	17.86	4	100.00	194	46.97
		CI	39	3.48	0	.00	42	10.17
		SSI	0	.00	0	.00	51	12.35
		SI	38	3.39	0	.00	0	.00
		C	22	1.96	0	.00	16	3.87
		CH	29	2.59	0	.00	5	1.21
		T	24	2.14	0	.00	6	1.45
		S	19	1.70	0	.00	1	.24
		SS	5	.45	0	.00	4	.97
		SC	4	.36	0	.00	2	.48
		SCI	4	.36	0	.00	1	.24
		X	3	.27	0	.00	0	.00
		CE	2	.18	0	.00	0	.00
		SCH	2	.18	0	.00	0	.00
/T/	7793		3756		526		3511	
		T	3722	99.09	526	100.00	3280	93.42
		TI	28	.74	0	.00	188	5.35
		ED	0	.00	0	.00	28	.80
		BT	0	.00	0	.00	11	.31
		TW	3	.08	0	.00	0	.00
		PT	1	.03	0	.00	1	.03
		CI	0	.00	0	.00	2	.06
		CHT	1	.03	0	.00	1	.03
		TH	1	.03	0	.00	0	.00
/T1/	411		226		0		185	
		TH	226	100.00	0	.00	185	100.00
/T2/	149		61		4		84	
		TH	61	100.00	4	100.00	84	100.00
/V/	1492		838		18		636	
		V	838	100.00	16	88.89	631	99.21
		LV	0	.00	2	11.11	2	.31
		F	0	.00	0	.00	3	.47

TABLE 18--Continued

Phoneme	Total Occurrence	Grapheme	Initial		Medial		Final	
			f	%	f	%	f	%
/W/	626		517		109		0	
		W	494	95.55	84	77.06	0	.00
		U	22	4.26	25	22.94	0	.00
		OU	1	.19	0	.00	0	.00
/Y/	120		118		2		0	
		I	64	54.24	2	100.00	0	.00
		Y	53	44.92	0	.00	0	.00
		E	1	.84	0	.00	0	.00
/Z/	995		310		96		589	
		S	168	54.19	96	100.00	376	63.84
		Z	84	27.10	0	.00	145	24.62
		ES	0	.00	0	.00	44	7.47
		X	42	13.55	0	.00	0	.00
		ZZ	1	.32	0	.00	22	3.74
		SS	11	3.55	0	.00	2	.34
		SC	3	.97	0	.00	0	.00
		CZ	1	.32	0	.00	0	.00
/ZH/	102		59		0		43	
		SI	35	59.32	0	.00	15	34.88
		S	19	32.20	0	.00	15	34.88
		G	3	5.08	0	.00	12	27.91
		Z	1	1.70	0	.00	1	2.33
		TI	1	1.70	0	.00	0	.00

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

/S/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

/SH/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

/Y/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

/Z/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

/ZH/: No graphemic option exceeds the 80% criterion in initial or final position in syllables.

Summarizing the effects of the position in syllables in which graphemic options for a particular phoneme occur, it is observed that graphemic options do not equally distribute themselves throughout a given position. Thus, when the distribution of graphemic options is tabulated for a given position, it is seen that in many instances a particular graphemic option will exceed the 80% criterion in either initial, medial, or final position in syllables.

What is the relevance of this observation for spelling instruction? Of considerable importance is the indication that the number of graphemic options used to represent a phoneme is not necessarily indicative of the complexity with which some phonemes are represented in the orthography. A consideration of only the number of different graphemic options per phoneme leads to the erroneous conclusion that all such options have an equal opportunity of representing a phoneme wherever it might occur in syllables and words.

To illustrate, consider /I3/ of the 22 vowel classification in Table 17, page 146. Twenty-two different graphemic options may represent this phoneme in the corpus. Yet, it was earlier noted that only two of these options, < I > and < Y >, represent /I3/ over 90% of the time. A tabulation of which graphemic option most often occurs in initial position in syllables reveals that < I > occurs in that position nearly 90% of the time, with the remaining 21 options occurring in initial position only 10% of the time in the corpus.

It is apparent that, when the distribution of graphemic options for a given position in syllables is tabulated, a number of options for each phoneme are rare spellings and quite possibly should be treated as such in elementary school spelling instruction.

Equally revealing is the observation that a number of phonemes having no single graphemic representation above the 80% criterion can be seen to exceed this criterion when distribution of graphemic options by position in syllables is tabulated. Yet, when the phonemes in a 22 vowel classification are examined for their approximations to the alphabetical principle irrespective of position and stress factors, 16 of them fall below the 80% criterion, with only /A3/, /A5/, /E3/, /O3/, /O5/, and /U3/ exceeding this criterion. When distribution of graphemic options for these phonemes by position in syllables is tabulated, nine of the 16 (/A/, /E/, /E5/, /I/, /I3/, /O/, /OI/, /OU/, and /U/) have at least one graphemic option which represents them in a given position in syllables over 80% of the time.

The 30 consonant phonemes, in turn, illustrate the effects of distribution of phoneme-grapheme correspondences by position in syllables, although less so than either of the vowel phoneme classifications. Nineteen consonant phonemes attain the 80% criterion when positional factors are not considered. But when distributions of graphemic options in initial and final positions in syllables are tabulated, 21 of the 30 consonant phonemes exceed the 80% criterion, leaving only nine which still fall below this figure, the phonemes /CH/, /J/, /NG/, /NL/, /S/, /SH/, /Y/, /Z/, and /ZH/.

The next factor to be considered, then, is the effect of syllabic stress upon the distribution of phoneme-grapheme correspondences among the respective positions in syllables in terms of achieving the 80% criterion. The interpretation to follow is concerned with determining whether any graphemic option, no matter how infrequent its occurrence, exceeds this criterion when position in stressed and unstressed syllables is considered.

Distribution of phoneme-grapheme correspondences by position in stressed and unstressed syllables: 22 vowel classification. As seen in Table 19; /A2/: No graphemic option exceeds the 80% criterion in any position in stressed and unstressed syllables.

/E2/: No graphemic option exceeds the 80% criterion in any position in stressed and unstressed syllables.

/02/: < AU > represents this phoneme 93.75% of the time in final position in unaccented syllables.

/06/: < OO > represents this phoneme 100% of the time in medial position in secondary accented syllables. < OO-E > is a single occurrence, found in initial position in a primary accented syllable, in the word ooze.

/07/: < U > represents this phoneme 83.09% of the time in medial position in unaccented syllables.

/ə/: No graphemic option exceeds the 80% criterion in any position in stressed and unstressed syllables, with the exception of one occurrence of < E > in medial position of a primary accented syllable in complimentary, and one occurrence of < A > in final position in a primary accented syllable in the article a.

/U2/: No graphemic option exceeds the 80% criterion in any position in stressed and unstressed syllables.

N.B. to Editor:

(The reason for the fact that Table 19 begins on page 162 rather than page 161 [immediately after the introduction of discussion of Table 19] is due to fact that Table 19 is on double facing pages so must be "even and odd" mmbered pages) thus forcing us to place Table on pages 162-3 and ff. This throws Table 20 even farther from opening of discussion starting on page 161. This matter may be improved when text is set in type).

30 consonant classification.[1] In Table 20: /CH/: < CH > represents this phoneme 98.64% of the time in initial position in primary accented syllables and 100% of the time in initial position in secondary accented syllables.

/J/: < G > represents this phoneme 82.46% of the time in initial position in unaccented syllables, and 94.57% of the time in final position in unaccented syllables.

/NG/: < NG > represents this phoneme 97.32% of the time in final position in unaccented syllables.

/NL/: No graphemic option exceeds the 80% criterion in any position in stressed and unstressed syllables.

/S/: < S > represents this phoneme 83.32% of the time in initial position in primary stressed syllables and 83.99% of the time in initial position in secondary stressed syllables.

/SH/: < TI > represents this phoneme 80.20% of the time in initial position in unaccented syllables. < SH > represents this phoneme 83.52% of the time in initial position in primary accented syllables and 89.66% of the time in initial position in secondary accented syllables. In final position, < SH > represents /SH/ 82.61% of the time in secondary accented syllables and 100% of the time in unaccented syllables.

/Y/: < I > represents this phoneme 86.30% of the time in initial position in unaccented syllables. < Y > represents this phoneme 97.50% of the time in initial position in primary accented syllables and 100% of the time in initial position in secondary accented syllables.

/Z/: No graphemic option exceeds the 80% criterion in initial or final position in stressed or unstressed syllables.

---

[1] See pages 176 through 185 for Table 20.

TABLE 19

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF THE DISTRIBUTION OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN STRESSED AND UNSTRESSED SYLLABLES: 22 VOWEL CLASSIFICATION

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/A/	2248		25		6		108	
		A	6	24.00	3	50.00	2	1.85
		A-E	13	52.00	3	50.00	105	97.22
		AI	4	16.00	0	.00	1	.93
		AY	0	.00	0	.00	0	.00
		AI-E	1	4.00	0	.00	0	.00
		EIGH	1	4.00	0	.00	0	.00
		E	0	.00	0	.00	0	.00
		EA	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
		EY	0	.00	0	.00	0	.00
		ET	0	.00	0	.00	0	.00
		E-E	0	.00	0	.00	0	.00
		AIGH	0	.00	0	.00	0	.00
		EI-E	0	.00	0	.00	0	.00
		AU-E	0	.00	0	.00	0	.00
		AY-E	0	.00	0	.00	0	.00
/A2/	220		18		0		0	
		A	1	5.56	0	.00	0	.00
		A-E	0	.00	0	.00	0	.00
		AI	9	50.00	0	.00	0	.00
		E-E	1	5.56	0	.00	0	.00
		EA	0	.00	0	.00	0	.00
		EI	3	16.67	0	.00	0	.00
		E	0	.00	0	.00	0	.00
		AI-E	3	16.67	0	.00	0	.00
		EY	1	5.56	0	.00	0	.00
/A3/	4340		336		104		620	
		A	336	100.00	104	100.00	569	91.77
		A-E	0	.00	0	.00	51	8.23
		AI	0	.00	0	.00	0	.00
/A5/	580		64		11		20	
		A	60	93.75	8	72.73	20	100.00
		A-E	1	1.56	0	.00	0	.00
		EA	0	.00	2	18.18	0	.00
		E	2	3.13	1	9.09	0	.00
		AH	0	.00	0	.00	0	.00
		AU	1	1.56	0	.00	0	.00

TABLE 19--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
634		82		331		916		112		34	
105	16.56	25	30.49	1	.30	778	84.93	71	63.39	11	32.35
315	49.68	31	37.80	323	97.58	0	.00	0	.00	0	.00
168	26.50	14	17.07	5	1.51	14	1.53	0	.00	2	5.88
0	.00	2	2.44	1	.30	81	8.84	40	35.71	7	20.59
16	2.52	0	.00	1	.30	0	.00	0	.00	0	.00
4	.63	3	3.66	0	.00	10	1.09	0	.00	0	.00
0	.00	0	.00	0	.00	5	.55	1	.89	10	29.41
7	1.10	6	7.32	0	.00	1	.11	0	.00	0	.00
11	1.74	0	.00	0	.00	3	.33	0	.00	0	.00
0	.00	0	.00	0	.00	14	1.53	0	.00	0	.00
0	.00	0	.00	0	.00	6	.66	0	.00	3	8.82
2	.32	0	.00	0	.00	3	.33	0	.00	1	2.94
3	.47	1	1.22	0	.00	0	.00	0	.00	0	.00
2	.32	0	.00	0	.00	0	.00	0	.00	0	.00
1	.16	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.11	0	.00	0	.00
153		26		23		0		0		0	
59	38.56	4	15.38	0	.00	0	.00	0	.00	0	.00
42	27.45	8	30.77	0	.00	0	.00	0	.00	0	.00
31	20.26	4	15.38	2	8.70	0	.00	0	.00	0	.00
7	4.58	8	30.77	18	78.26	0	.00	0	.00	0	.00
11	7.19	2	7.69	0	.00	0	.00	0	.00	0	.00
2	1.31	0	.00	0	.00	0	.00	0	.00	0	.00
1	.65	0	.00	3	13.04	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
1846		239		891		0		1		303	
1808	97.94	237	99.16	834	93.60	0	.00	1	100.00	303	100.00
37	2.00	2	.84	57	6.40	0	.00	0	.00	0	.00
1	.05	0	.00	0	.00	0	.00	0	.00	0	.00
330		34		60		49		3		9	
284	86.06	31	91.18	57	95.00	46	93.88	3	100.00	9	100.00
26	7.88	2	5.88	2	3.33	0	.00	0	.00	0	.00
15	4.55	1	2.94	0	.00	0	.00	0	.00	0	.00
1	.30	0	.00	1	1.67	0	.00	0	.00	0	.00
1	.30	0	.00	0	.00	3	6.12	0	.00	0	.00
3	.91	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 19--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/E/	2538		22		0		3	
		E	0	.00	0	.00	1	33.33
		EE	1	4.55	0	.00	0	.00
		EA	16	72.73	0	.00	1	33.33
		E-E	3	13.64	0	.00	0	.00
		I-E	1	4.55	0	.00	0	.00
		I	0	.00	0	.00	1	33.33
		IE	0	.00	0	.00	0	.00
		EA-E	1	4.55	0	.00	0	.00
		IE-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
		EE-E	0	.00	0	.00	0	.00
		EI-E	0	.00	0	.00	0	.00
		EY	0	.00	0	.00	0	.00
		AE	0	.00	0	.00	0	.00
		OE	0	.00	0	.00	0	.00
		EO	0	.00	0	.00	0	.00
/E2/	198		8		0		0	
		E	1	12.50	0	.00	0	.00
		EA	4	50.00	0	.00	0	.00
		EE	2	25.00	0	.00	0	.00
		E-E	0	.00	0	.00	0	.00
		IE	1	12.50	0	.00	0	.00
		I	0	.00	0	.00	0	.00
		IE-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
/E3/	3646		222		94		482	
		E	212	95.50	64	68.09	473	98.13
		EA	0	.00	0	.00	0	.00
		A	2	.90	30	31.91	0	.00
		E-E	8	3.60	0	.00	7	1.45
		AI	0	.00	0	.00	0	.00
		IE	0	.00	0	.00	0	.00
		EO	0	.00	0	.00	0	.00
		U	0	.00	0	.00	0	.00
		A-E	0	.00	0	.00	1	.21
		AE	0	.00	0	.00	1	.21
		AY	0	.00	0	.00	0	.00
		EA-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00

TABLE 19--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
521		28		43		413		93		1415	
18	3.45	0	.00	6	13.95	258	62.47	87	93.55	1395	98.59
153	29.37	11	39.29	2	4.65	69	16.71	3	3.23	10	.71
167	32.05	8	28.57	0	.00	53	12.83	0	.00	0	.00
42	8.06	1	3.57	16	37.21	0	.00	0	.00	0	.00
28	5.37	1	3.57	14	32.56	0	.00	0	.00	0	.00
13	2.50	0	.00	1	2.33	15	3.63	1	1.08	7	.50
30	5.76	2	7.14	1	2.33	0	.00	0	.00	0	.00
27	5.18	0	.00	2	4.65	0	.00	0	.00	0	.00
17	3.26	5	17.86	1	2.33	0	.00	0	.00	0	.00
11	2.11	0	.00	0	.00	5	1.21	0	.00	0	.00
9	1.73	0	.00	0	.00	0	.00	0	.00	0	.00
6	1.15	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	5	1.21	1	1.08	0	.00
0	.00	0	.00	0	.00	2	.48	0	.00	3	.21
0	.00	0	.00	0	.00	5	1.21	0	.00	0	.00
0	.00	0	.00	0	.00	1	.24	1	1.08	0	.00
165		13		12		0		0		0	
53	32.12	9	69.23	1	8.33	0	.00	0	.00	0	.00
44	26.67	1	7.69	0	.00	0	.00	0	.00	0	.00
32	19.39	1	7.69	1	8.33	0	.00	0	.00	0	.00
15	9.09	2	15.38	10	83.33	0	.00	0	.00	0	.00
13	7.88	0	.00	0	.00	0	.00	0	.00	0	.00
3	1.82	0	.00	0	.00	0	.00	0	.00	0	.00
3	1.82	0	.00	0	.00	0	.00	0	.00	0	.00
2	1.21	0	.00	0	.00	0	.00	0	.00	0	.00
2056		353		395		1		0		43	
1868	90.86	275	77.90	380	96.20	1	100.00	0	.00	43	100.00
119	5.79	14	3.97	6	1.52	0	.00	0	.00	0	.00
1	.05	58	16.43	3	.76	0	.00	0	.00	0	.00
53	2.58	5	1.52	0	.00	0	.00	0	.00	0	.00
3	.15	1	.28	0	.00	0	.00	0	.00	0	.00
4	.19	0	.00	0	.00	0	.00	0	.00	0	.00
3	.15	0	.00	0	.00	0	.00	0	.00	0	.00
2	.10	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	.05	0	.00	0	.00	0	.00	0	.00	0	.00
1	.05	0	.00	0	.00	0	.00	0	.00	0	.00
1	.05	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 19--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/E5/	2170		0		0		963	
		E	0	.00	0	.00	820	85.15
		O	0	.00	0	.00	70	7.27
		A	0	.00	0	.00	52	5.40
		U	0	.00	0	.00	8	.83
		U-E	0	.00	0	.00	8	.83
		I	0	.00	0	.00	1	.10
		Y	0	.00	0	.00	2	.21
		E-E	0	.00	0	.00	1	.10
		OU	0	.00	0	.00	1	.10
/I/	1482							
			9		0		46	
		I-E	6	66.67	0	.00	46	100.00
		I	2	22.22	0	.00	0	.00
		Y	0	.00	0	.00	0	.00
		IGH	0	.00	0	.00	0	.00
		IE	0	.00	0	.00	0	.00
		Y-E	0	.00	0	.00	0	.00
		EY-E	0	.00	0	.00	0	.00
		EI	0	.00	0	.00	0	.00
		AY	0	.00	0	.00	0	.00
		EIGH	0	.00	0	.00	0	.00
		UY	0	.00	0	.00	0	.00
		AI-E	1	11.11	0	.00	0	.00
		EY	0	.00	0	.00	0	.00
		OY	0	.00	0	.00	0	.00
/I3/	7815		195		200		1322	
		I	192	98.46	200	100.00	1143	86.46
		Y	0	.00	0	.00	5	.38
		I-E	0	.00	0	.00	44	3.33
		A-E	0	.00	0	.00	108	8.17
		EY	0	.00	0	.00	0	.00
		IE	0	.00	0	.00	5	.38
		UI	0	.00	0	.00	0	.00
		AI	0	.00	0	.00	1	.08
		EI	3	1.54	0	.00	5	.38
		EE	0	.00	0	.00	0	.00
		E	0	.00	0	.00	0	.00
		A	0	.00	0	.00	4	.30
		IA-E	0	.00	0	.00	3	.23
		U	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	3	.23
		AY	0	.00	0	.00	0	.00
		E-E	0	.00	0	.00	0	.00

TABLE 19--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
0		0		1201		1		0		5	
0	.00	0	.00	840	69.94	1	100.00	0	.00	5	100.00
0	.00	0	.00	198	16.49	0	.00	0	.00	0	.00
0	.00	0	.00	116	9.66	0	.00	0	.00	0	.00
0	.00	0	.00	23	1.92	0	.00	0	.00	0	.00
0	.00	0	.00	15	1.25	0	.00	0	.00	0	.00
0	.00	0	.00	7	.58	0	.00	0	.00	0	.00
0	.00	0	.00	2	.17	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
531		103		147		389		68		189	
313	58.95	61	59.22	129	87.76	0	.00	0	.00	0	.00
140	26.37	16	15.53	1	.68	253	65.04	41	60.29	101	53.44
2	.38	0	.00	0	.00	97	24.94	26	38.24	86	45.50
53	9.98	24	23.30	1	.68	10	2.57	0	.00	0	.00
8	1.51	0	.00	6	4.08	10	2.57	1	1.47	1	.53
8	1.51	1	.97	10	6.80	4	1.03	0	.00	0	.00
0	.00	0	.00	0	.00	7	1.80	0	.00	0	.00
4	.75	1	.97	0	.00	1	.26	0	.00	0	.00
0	.00	0	.00	0	.00	2	.51	0	.00	1	.53
3	.56	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	3	.77	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.26	0	.00	0	.00
0	.00	0	.00	0	.00	1	.26	0	.00	0	.00
1957		242		1147		0		0		2752	
1801	92.03	224	92.56	747	65.13	0	.00	0	.00	1039	37.75
101	5.16	16	6.61	32	2.79	0	.00	0	.00	1647	59.84
35	1.79	1	.41	259	22.58	0	.00	0	.00	0	.00
1	.05	0	.00	78	6.80	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	40	1.45
0	.00	0	.00	8	.70	0	.00	0	.00	16	.58
10	.51	1	.41	4	.35	0	.00	0	.00	1	.04
0	.00	0	.00	14	1.22	0	.00	0	.00	0	.00
0	.00	0	.00	3	.26	0	.00	0	.00	0	.00
2	.10	0	.00	0	.00	0	.00	0	.00	4	.15
2	.10	0	.00	0	.00	0	.00	0	.00	3	.11
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
3	.15	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	.04
0	.00	0	.00	1	.09	0	.00	0	.00	0	.00

TABLE 19--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/I3/--Continued								
		EA	0	.00	0	.00	0	.00
		EA-E	0	.00	0	.00	1	.08
		IE-E	0	.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		Y-E	0	.00	0	.00	0	.00
/O/	2587		17		1		14	
		O	2	11.76	0	.00	3	21.43
		O-E	2	11.76	0	.00	9	64.29
		OA	8	47.06	0	.00	0	.00
		OW	4	23.53	1	100.00	2	14.29
		OU	0	.00	0	.00	0	.00
		OE	0	.00	0	.00	0	.00
		OU-E	0	.00	0	.00	0	.00
		OO	0	.00	0	.00	0	.00
		OUGH	0	.00	0	.00	0	.00
		EAU	0	.00	0	.00	0	.00
		OH	1	5.88	0	.00	0	.00
		AU	0	.00	0	.00	0	.00
		EW	0	.00	0	.00	0	.00
		OA-E	0	.00	0	.00	0	.00
		AU-E	0	.00	0	.00	0	.00
		OW-E	0	.00	0	.00	0	.00
/OI/	149		5		0		3	
		OI	5	100.00	0	.00	3	100.00
		OY	0	.00	0	.00	0	.00
		OI-E	0	.00	0	.00	0	.00
		OY-E	0	.00	0	.00	0	.00
/OU/	406		45		8		19	
		OU	42	93.33	8	100.00	19	100.00
		OW	2	4.44	0	.00	0	.00
		OU-E	1	2.22	0	.00	0	.00
		OUGH	0	.00	0	.00	0	.00
		OW-E	0	.00	0	.00	0	.00
/O2/	767		52		12		21	
		O	31	59.62	6	50.00	9	42.86
		A	15	28.85	5	41.67	6	28.57
		AU	2	3.85	1	8.33	6	28.57
		AW	2	3.85	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00

TABLE 19--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
0	.00	0	.00	0	.00	0	.00	0	.00	1	.04
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
1	.05	0	.00	0	.00	0	.00	0	.00	0	.00
1	.05	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	1	.09	0	.00	0	.00	0	.00
561		108		115		482		141		1148	
180	32.09	26	24.07	36	31.30	413	85.68	132	93.62	1084	94.43
244	43.49	42	38.89	73	63.48	0	.00	0	.00	0	.00
80	14.26	34	31.48	2	1.74	1	.21	0	.00	1	.09
12	2.14	2	1.85	0	.00	46	9.54	6	4.26	51	4.44
27	4.81	1	.93	1	.87	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	10	2.07	1	.71	2	.17
6	1.07	1	.93	3	2.61	0	.00	0	.00	0	.00
7	1.25	2	1.85	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	4	.83	0	.00	4	.35
0	.00	0	.00	0	.00	3	.62	0	.00	3	.26
0	.00	0	.00	0	.00	1	.21	2	1.42	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	3	.26
0	.00	0	.00	0	.00	3	.62	0	.00	0	.00
3	.53	0	.00	0	.00	0	.00	0	.00	0	.00
2	.36	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.21	0	.00	0	.00
56		6		18		53		2		6	
50	89.29	6	100.00	15	83.33	12	22.64	0	.00	1	16.67
0	.00	0	.00	0	.00	41	77.36	2	100.00	5	83.33
6	10.71	0	.00	2	11.11	0	.00	0	.00	0	.00
0	.00	0	.00	1	5.56	0	.00	0	.00	0	.00
200		47		7		69		9		2	
125	62.50	21	44.68	7	100.00	5	7.25	0	.00	0	.00
40	20.00	5	10.64	0	.00	61	88.41	9	100.00	2	100.00
32	16.00	21	44.68	0	.00	0	.00	0	.00	0	.00
1	.50	0	.00	0	.00	3	4.35	0	.00	0	.00
2	1.00	0	.00	0	.00	0	.00	0	.00	0	.00
435		41		54		116		20		16	
204	46.90	21	51.22	41	75.93	0	.00	0	.00	0	.00
102	23.45	17	41.46	6	11.11	12	10.34	2	10.00	0	.00
51	11.72	0	.00	6	11.11	53	45.69	12	60.00	15	93.75
21	4.83	0	.00	1	1.85	45	38.79	5	25.00	1	6.25
17	3.91	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 19--Continued

			Initial					
	Total		Primary		Secondary		Unaccented	
Pho-neme	Occur-rence	Graph-eme	f	%	f	%	f	%
/02/--Continued								
		OUGH	1	1.92	0	.00	0	.00
		AUGH	1	1.92	0	.00	0	.00
		OA	0	.00	0	.00	0	.00
		AU-E	0	.00	0	.00	0	.00
		A-E	0	.00	0	.00	0	.00
		AO	0	.00	0	.00	0	.00
		AW-E	0	.00	0	.00	0	.00
/03/	1662		158		30		26	
		O	158	100.00	30	100.00	26	100.00
		A	0	.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		OW	0	.00	0	.00	0	.00
/05/	127		10		1		0	
		O	10	100.00	1	100.00	0	.00
		O-E	0	.00	0	.00	0	.00
/06/	453		1		0		0	
		OO	0	.00	0	.00	0	.00
		U	0	.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	0	.00
		OU	0	.00	0	.00	0	.00
		EW	0	.00	0	.00	0	.00
		UE	0	.00	0	.00	0	.00
		OO-E	1	100.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		UI	0	.00	0	.00	0	.00
		EU	0	.00	0	.00	0	.00
		OE	0	.00	0	.00	0	.00
		UI-E	0	.00	0	.00	0	.00
		OU-E	0	.00	0	.00	0	.00
		OUGH	0	.00	0	.00	0	.00
		UO	0	.00	0	.00	0	.00
/07/	368		0		0		0	
		U	0	.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		OU	0	.00	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	0	.00
		EU	0	.00	0	.00	0	.00

TABLE 19--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
12	2.76	2	4.88	0	.00	0	.00	0	.00	0	.00
5	1.15	1	2.44	0	.00	4	3.45	1	5.00	0	.00
9	2.07	0	.00	0	.00	0	.00	0	.00	0	.00
9	2.07	0	.00	0	.00	0	.00	0	.00	0	.00
3	.69	0	.00	0	.00	0	.00	0	.00	0	.00
2	.46	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	2	1.72	0	.00	0	.00
1151		187		109		0		0		1	
1056	91.75	181	96.79	106	97.25	0	.00	0	.00	1	100.00
71	6.17	6	3.21	3	2.75	0	.00	0	.00	0	.00
20	1.74	0	.00	0	.00	0	.00	0	.00	0	.00
4	.35	0	.00	0	.00	0	.00	0	.00	0	.00
81		14		21		0		0		0	
78	96.30	13	92.86	21	100.00	0	.00	0	.00	0	.00
3	3.70	1	7.14	0	.00	0	.00	0	.00	0	.00
231		22		4		158		11		26	
123	53.25	22	100.00	3	75.00	21	13.29	2	18.18	2	7.69
11	4.76	0	.00	0	.00	67	42.41	2	18.18	13	50.00
16	6.93	0	.00	0	.00	16	10.13	1	9.09	4	15.38
33	14.29	0	.00	1	25.00	0	.00	0	.00	0	.00
9	3.90	0	.00	0	.00	13	8.23	2	18.18	5	19.23
2	.87	0	.00	0	.00	19	12.03	1	9.09	0	.00
0	.00	0	.00	0	.00	16	10.13	0	.00	0	.00
11	4.76	0	.00	0	.00	0	.00	0	.00	0	.00
12	5.20	0	.00	0	.00	0	.00	0	.00	0	.00
6	2.60	0	.00	0	.00	0	.00	0	.00	0	.00
1	.43	0	.00	0	.00	2	1.27	0	.00	1	3.85
0	.00	0	.00	0	.00	1	.63	3	27.27	0	.00
4	1.73	0	.00	0	.00	0	.00	0	.00	0	.00
3	1.30	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.63	0	.00	1	3.85
0	.00	0	.00	0	.00	2	1.27	0	.00	0	.00
155		31		136		0		0		46	
46	29.68	5	16.13	113	83.09	0	.00	0	.00	36	78.26
75	48.39	23	74.19	16	11.76	0	.00	0	.00	0	.00
15	9.68	0	.00	6	4.41	0	.00	0	.00	4	8.70
8	5.16	3	9.68	6	4.00	0	.00	0	.00	6	13.04
10	6.45	0	.00	1	.74	0	.00	0	.00	0	.00
1	.65	0	.00	0	. .00	0	.00	0	.00	0	.00

TABLE 19--Continued

	Total		Initial					
Pho-	Occur-	Graph-	Primary		Secondary		Unaccented	
neme	rence	eme	f	%	f	%	f	%
/ə/	6013		0		0		911	
		O	0	.00	0	.00	321	35.24
		A	0	.00	0	.00	6	.66
		I	0	.00	0	.00	6	.66
		E	0	.00	0	.00	228	25.03
		OU	0	.00	0	.00	191	20.97
		U	0	.00	0	.00	90	9.88
		E-E	0	.00	0	.00	37	4.06
		Y	0	.00	0	.00	2	.22
		IE	0	.00	0	.00	13	1.43
		O-E	0	.00	0	.00	1	.11
		EO	0	.00	0	.00	5	.55
		AI	0	.00	0	.00	3	.33
		EOU	0	.00	0	.00	0	.00
		I-E	0	.00	0	.00	2	.22
		IE-E	0	.00	0	.00	1	.11
		U-E	0	.00	0	.00	1	.11
		A-E	0	.00	0	.00	1	.11
		EA	0	.00	0	.00	2	.22
		EI	0	.00	0	.00	1	.11
		IA	0	.00	0	.00	0	.00
		OI-E	0	.00	0	.00	0	.00
		AU	0	.00	0	.00	0	.00
/U/	1188		15		0		26	
		U	7	46.67	0	.00	2	7.69
		U-E	8	53.33	0	.00	23	88.46
		EW	0	.00	0	.00	0	.00
		EU	0	.00	0	.00	0	.00
		UE	0	.00	0	.00	1	3.85
		UI	0	.00	0	.00	0	.00
		IEW	0	.00	0	.00	0	.00
		EAU	0	.00	0	.00	0	.00
		IEU	0	.00	0	.00	0	.00
		EU-E	0	.00	0	.00	0	.00
		EW-E	0	.00	0	.00	0	.00
/U2/	787		35		0		3	
		E	9	25.71	0	.00	1	33.33
		U	7	20.00	0	.00	1	33.33
		I	2	5.71	0	.00	0	.00
		O	0	.00	0	.00	0	.00
		E-E	1	2.86	0	.00	1	33.33
		EA	15	42.86	0	.00	0	.00

TABLE 19--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
1		0		2045		1		0		3055	
0	.00	0	.00	1176	57.51	0	.00	0	.00	114	3.73
0	.00	0	.00	13	.64	1	100.00	0	.00	1418	46.42
0	.00	0	.00	9	.44	0	.00	0	.00	1332	43.60
1	100.00	0	.00	419	20.49	0	.00	0	.00	115	3.76
0	.00	0	.00	144	7.04	0	.00	0	.00	1	.03
0	.00	0	.00	165	8.07	0	.00	0	.00	42	1.37
0	.00	0	.00	64	3.13	0	.00	0	.00	0	.00
0	.00	0	.00	1	.05	0	.00	0	.00	20	.65
0	.00	0	.00	7	.34	0	.00	0	.00	2	.07
0	.00	0	.00	20	.98	0	.00	0	.00	0	.00
0	.00	0	.00	5	.25	0	.00	0	.00	0	.00
0	.00	0	.00	2	.10	0	.00	0	.00	4	.13
0	.00	0	.00	8	.39	0	.00	0	.00	0	.00
0	.00	0	.00	4	.20	0	.00	0	.00	0	.00
0	.00	0	.00	3	.15	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	3	.10
0	.00	0	.00	2	.10	0	.00	0	.00	0	.00
0	.00	0	.00	1	.05	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	.03
0	.00	0	.00	0	.00	0	.00	0	.00	2	.07
0	.00	0	.00	2	.10	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	1	.03
138		1		138		326		61		483	
22	15.94	0	.00	13	9.42	265	81.29	55	90.16	450	93.17
101	73.19	0	.00	124	89.86	0	.00	0	.00	0	.00
7	5.07	0	.00	0	.00	24	7.36	1	1.64	6	1.24
1	.72	0	.00	1	.72	13	3.99	4	6.56	9	1.86
0	.00	0	.00	0	.00	11	3.37	0	.00	15	3.11
6	4.35	1	100.00	0	.00	1	.31	0	.00	0	.00
0	.00	0	.00	0	.00	4	1.23	1	1.64	1	.21
0	.00	0	.00	0	.00	4	1.23	0	.00	1	.21
0	.00	0	.00	0	.00	3	.92	0	.00	1	.21
1	.72	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.31	0	.00	0	.00
620		86		43		0		0		0	
244	39.35	35	40.70	24	55.81	0	.00	0	.00	0	.00
171	27.58	10	11.63	14	32.56	0	.00	0	.00	0	.00
83	13.39	18	20.93	1	2.33	0	.00	0	.00	0	.00
31	5.00	22	25.58	0	.00	0	.00	0	.00	0	.00
36	5.81	0	.00	3	6.98	0	.00	0	.00	0	.00
14	2.26	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 19--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/U2/--Continued								
		OU	0	.00	0	.00	0	.00
		U-E	1	2.86	0	.00	0	.00
		EU	0	.00	0	.00	0	.00
		EA-E	0	.00	0	.00	0	.00
		Y	0	.00	0	.00	0	.00
		I-E	0	.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		OU-E	0	.00	0	.00	0	.00
/U3/	1410		63		49		163	
		U	58	92.06	49	100.00	163	100.00
		O	5	7.94	0	.00	0	.00
		OU	0	.00	0	.00	0	.00
		O-E	0	.00	0	.00	0	.00
		U-E	0	.00	0	.00	0	.00
		OO	0	.00	0	.00	0	.00

---

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

TABLE 19--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
20	3.23	1	1.16	0	.00	0	.00	0	.00	0	.00
8	1.29	0	.00	1	2.33	0	.00	0	.00	0	.00
6	.97	0	.00	0	.00	0	.00	0	.00	0	.00
2	.32	0	.00	0	.00	0	.00	0	.00	0	.00
2	.32	0	.00	0	.00	0	.00	0	.00	0	.00
1	.16	0	.00	0	.00	0	.00	0	.00	0	.00
1	.16	0	.00	0	.00	0	.00	0	.00	0	.00
1	.16	0	.00	0	.00	0	.00	0	.00	0	.00
999		78		56		0		0		2	
822	82.28	67	85.90	51	91.07	0	.00	0	.00	2	100.00
95	9.51	10	12.82	2	3.57	0	.00	0	.00	0	.00
29	2.90	0	.00	1	1.79	0	.00	0	.00	0	.00
24	2.40	1	1.28	1	1.79	0	.00	0	.00	0	.00
22	2.20	0	.00	1	1.79	0	.00	0	.00	0	.00
7	.70	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 20

FREQUENCY AND PERCENTAGE[a] TABULATIONS OF THE DISTRIBUTION OF PHONEME-GRAPHEME CORRESPONDENCES IN AMERICAN-ENGLISH BY POSITION IN STRESSED AND UNSTRESSED SYLLABLES: 30 CONSONANT CLASSIFICATION

Pho-neme	Total Occur-rence	Graph-eme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/B/	2303		967		158		728	
		B	964	99.69	155	98.10	728	100.00
		BB	3	.31	2	1.27	0	.00
		PB	0	.00	1	.63	0	.00
/CH/	564		147		6		166	
		CH	145	98.64	6	100.00	26	15.66
		T	0	.00	0	.00	127	76.51
		TCH	0	.00	0	.00	0	.00
		TI	0	.00	0	.00	13	7.83
		C	2	1.36	0	.00	0	.00
/D/	3691		802		163		1141	
		D	795	99.13	161	98.77	1141	100.00
		DD	7	.87	0	.00	0	.00
		LD	0	.00	2	1.23	0	.00
/F/	2019		1007		107		516	
		F	889	88.28	89	83.18	403	78.10
		PH	59	5.86	14	13.08	109	21.12
		FF	59	5.86	2	1.87	1	.19
		LF	0	.00	2	1.87	3	.58
		GH	0	.00	0	.00	0	.00
		FT	0	.00	0	.00	0	.00
/G/	1338		565		51		289	
		G	538	95.22	47	92.16	286	98.96
		GG	5	.89	0	.00	0	.00
		X	0	.00	0	.00	0	.00
		GUE	0	.00	0	.00	0	.00
		GU	15	2.65	4	7.84	0	.00
		GH	7	1.24	0	.00	2	.69
		TG	0	.00	0	.00	1	.35
/H/	778		585		90		103	
		H	573	97.95	87	96.67	102	99.03
		WH	12	2.05	3	3.33	1	.97
/HW/	89		63		12		14	
		WH	63	100.00	12	100.00	14	100.00

TABLE 20--*Continued*

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
1		0		0		294		48		107	
1	100.00	0	.00	0	.00	238	80.95	46	95.83	107	100.00
0	.00	0	.00	0	.00	56	19.05	2	4.17	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
2		0		0		218		16		9	
2	100.00	0	.00	0	.00	124	56.88	4	25.00	6	66.67
0	.00	0	.00	0	.00	35	16.06	10	62.50	3	33.33
0	.00	0	.00	0	.00	59	27.06	2	12.50	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
17		1		12		961		172		422	
16	94.12	1	100.00	12	100.00	892	92.82	171	99.42	422	100.00
1	5.88	0	.00	0	.00	65	6.76	1	.58	0	.00
0	.00	0	.00	0	.00	4	.42	0	.00	0	.00
41		6		5		275		36		26	
37	90.24	5	83.33	1	20.00	138	50.18	16	44.44	2	7.69
4	9.76	1	16.67	4	80.00	32	11.64	4	11.11	15	57.69
0	.00	0	.00	0	.00	90	32.73	16	44.44	9	34.62
0	.00	0	.00	0	.00	4	1.45	0	.00	0	.00
0	.00	0	.00	0	.00	8	2.91	0	.00	0	.00
0	.00	0	.00	0	.00	3	1.09	0	.00	0	.00
1		0		0		298		38		96	
1	100.00	0	.00	0	.00	228	76.51	33	86.84	45	46.88
0	.00	0	.00	0	.00	59	19.80	3	7.90	0	.00
0	.00	0	.00	0	.00	0	.00	2	5.26	40	41.67
0	.00	0	.00	0	.00	10	3.36	0	.00	11	11.45
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.33	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0		0		0		0		0		0	
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0		0		0		0		0		0	
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 20--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/H9/	41		16		2		3	
		H	16	100.00	2	100.00	3	100.00
/J/	982		296		23		268	
		G	104	35.14	15	65.22	221	82.46
		J	178	60.14	8	34.78	28	10.45
		DG	0	.00	0	.00	0	.00
		D	0	.00	0	.00	7	2.61
		DJ	13	4.39	0	.00	0	.00
		GI	0	.00	0	.00	11	4.10
		GG	0	.00	0	.00	0	.00
		DI	1	.33	0	.00	1	.37
/K/	4712		1365		191		1051	
		C	1204	88.20	169	88.48	914	86.97
		K	70	5.13	10	5.24	82	7.80
		CK	0	.00	0	.00	0	.00
		CH	43	3.15	5	2.62	48	4.57
		X	0	.00	0	.00	0	.00
		CC	41	3.00	4	2.09	1	.09
		QU	5	.37	0	.00	3	.29
		Q	0	.00	0	.00	2	.19
		LK	0	.00	0	.00	0	.00
		CQ	0	.00	0	.00	0	.00
		KH	2	.15	0	.00	1	.09
		SC	0	.00	3	1.57	0	.00
		CCH	0	.00	0	.00	0	.00
/KS/	271		0		0		0	
		X	0	.00	0	.00	0	.00
		CS	0	.00	0	.00	0	.00
/KW/	196		112		6		52	
		QU	107	95.54	6	100.00	52	100.00
		CQU	5	4.46	0	.00	0	.00
/L/	5389		957		127		792	
		L	875	91.43	118	92.91	776	97.98
		LL	82	8.57	9	7.09	13	1.64
		SL	0	.00	0	.00	3	.38
		LN	0	.00	0	.00	0	.00
/Ll/	651		0		0		0	
		LE	0	.00	0	.00	0	.00
		EL	0	.00	0	.00	0	.00

TABLE 20--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
8		2		1		0		0		9	
8	100.00	2	100.00	1	100.00	0	.00	0	.00	9	100.00
0		0		0		243		23		129	
0	.00	0	.00	0	.00	169	69.55	16	69.57	122	94.57
0	.00	0	.00	0	.00	3	1.23	1	4.35	0	.00
0	.00	0	.00	0	.00	44	18.11	1	4.35	6	4.65
0	.00	0	.00	0	.00	20	8.23	4	17.39	1	.78
0	.00	0	.00	0	.00	3	1.23	0	.00	0	.00
0	.00	0	.00	0	.00	3	1.23	0	.00	0	.00
0	.00	0	.00	0	.00	1	.41	1	4.35	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
263		12		72		1105		189		464	
213	80.99	4	33.33	69	95.83	414	37.47	57	30.16	408	87.93
40	15.21	7	58.33	0	.00	320	28.96	58	30.69	14	3.02
1	.38	0	.00	0	.00	228	20.63	35	18.52	26	5.60
9	3.42	1	8.33	3	4.17	13	1.18	9	4.76	11	2.37
0	.00	0	.00	0	.00	60	5.43	15	7.94	5	1.08
0	.00	0	.00	0	.00	23	2.08	7	3.70	0	.00
0	.00	0	.00	0	.00	19	1.72	0	.00	0	.00
0	.00	0	.00	0	.00	16	1.45	2	1.06	0	.00
0	.00	0	.00	0	.00	11	.99	3	1.59	0	.00
0	.00	0	.00	0	.00	0	.00	3	1.58	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.09	0	.00	0	.00
6		0		2		63		22		178	
6	100.00	0	.00	2	100.00	63	100.00	22	100.00	152	85.39
0	.00	0	.00	0	.00	0	.00	0	.00	26	14.61
26		0		0		0		0		0	
26	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
885		64		160		1245		164		995	
885	100.00	63	98.44	160	100.00	919	73.82	129	78.66	969	97.39
0	.00	1	1.56	0	.00	323	25.94	35	21.34	26	2.61
0	.00	0	.00	0	.00	2	.16	0	.00	0	.00
0	.00	0	.00	0	.00	1	.08	0	.00	0	.00
0		0		1		0		0		650	
0	.00	0	.00	1	100.00	0	.00	0	.00	619	95.23
0	.00	0	.00	0	.00	0	.00	0	.00	19	2.92

TABLE 20--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/Ll/--Continued								
		IL	0	.00	0	.00	0	.00
		AL	0	.00	0	.00	0	.00
		OL	0	.00	0	.00	0	.00
/M/	3501		1009		166		767	
		M	956	94.75	160	96.39	758	98.83
		MM	53	5.25	5	3.01	3	.39
		MB	0	.00	0	.00	0	.00
		LM	0	.00	0	.00	2	.26
		GM	0	.00	1	.60	4	.52
		MN	0	.00	0	.00	0	.00
/Ml/	97		0		0		0	
		M	0	.00	0	.00	0	.00
/N/	7656		605		70		628	
		N	539	89.09	64	91.43	620	98.73
		NN	25	4.13	3	4.29	2	.32
		KN	36	5.95	2	2.86	3	.48
		GN	5	.83	0	.00	0	.00
		PN	0	.00	1	1.42	2	.32
		MN	0	.00	0	.00	1	.15
/NG/	615		0		0		0	
		NG	0	.00	0	.00	0	.00
		N	0	.00	0	.00	0	.00
		NGUE	0	.00	0	.00	0	.00
/Nl/	128		0		0		0	
		EN	0	.00	0	.00	0	.00
		ON	0	.00	0	.00	0	.00
		IN	0	.00	0	.00	0	.00
/P/	3449		1412		201		755	
		P	1350	95.61	197	98.01	755	100.00
		PP	62	4.39	4	1.99	0	.00
/R/	9390		766		141		855	
		R	661	86.29	135	95.74	852	99.65
		RR	50	6.53	0	.00	0	.00
		WR	43	5.61	5	3.55	0	.00
		RH	12	1.57	1	.71	3	.35

TABLE 20--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
0	.00	0	.00	0	.00	0	.00	0	.00	7	1.08
0	.00	0	.00	0	.00	0	.00	0	.00	4	.62
0	.00	0	.00	0	.00	0	.00	0	.00	1	.15
113		4		4		836		146		456	
111	98.23	4	100.00	4	100.00	727	86.96	133	91.10	449	98.46
0	.00	0	.00	0	.00	67	8.01	11	7.53	1	.22
0	.00	0	.00	0	.00	24	2.87	2	1.37	1	.22
2	1.77	0	.00	0	.00	13	1.56	0	.00	0	.00
0	.00	0	.00	0	.00	1	.12	0	.00	2	.44
0	.00	0	.00	0	.00	4	.48	0	.00	3	.66
0		0		0		6		0		91	
0	.00	0	.00	0	.00	6	100.00	0	.00	91	100.00
499		52		817		1897		442		2646	
499	100.00	52	100.00	817	100.00	1785	94.10	436	98.64	2640	99.77
0	.00	0	.00	0	.00	90	4.74	6	1.36	1	.04
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	22	1.16	0	.00	5	.19
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
121		2		11		197		23		261	
5	4.13	0	.00	0	.00	92	46.70	11	47.83	254	97.32
116	95.87	2	100.00	11	100.00	103	52.29	12	52.17	7	2.68
0	.00	0	.00	0	.00	2	1.01	0	.00	0	.00
0		0		0		0		0		128	
0	.00	0	.00	0	.00	0	.00	0	.00	84	65.63
0	.00	0	.00	0	.00	0	.00	0	.00	41	32.03
0	.00	0	.00	0	.00	0	.00	0	.00	3	2.34
225		15		38		599		104		100	
224	99.56	15	100.00	38	100.00	534	89.15	84	80.77	99	99.00
1	.44	0	.00	0	.00	65	10.85	20	19.23	1	1.00
2071		244		860		1745		317		2391	
2071	100.00	244	100.00	860	100.00	1612	92.38	294	92.74	2390	99.96
0	.00	0	.00	0	.00	133	7.68	23	7.26	1	.04
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00

TABLE 20--Continued

Pho-neme	Total Occur-rence	Graph-eme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/S/	6326		2056		281		1047	
		S	1713	83.32	236	83.99	685	65.43
		C	244	11.87	24	8.54	269	25.69
		SS	53	2.58	12	4.27	5	.48
		SC	32	1.56	3	1.07	4	.38
		X	4	.19	0	.00	73	6.97
		ST	0	.00	0	.00	0	.00
		PS	8	.39	5	1.78	6	.57
		Z	0	.00	0	.00	2	.19
		SW	1	.05	1	.35	2	.19
		SCH	1	.05	0	.00	1	.10
/SH/	1537		182		29		909	
		TI	0	.00	0	.00	729	80.20
		SH	152	83.52	26	89.66	22	2.42
		CI	0	.00	0	.00	39	4.29
		SSI	0	.00	0	.00	0	.00
		SI	0	.00	0	.00	38	4.18
		C	0	.00	0	.00	22	2.42
		CH	15	8.24	3	10.34	11	1.21
		T	0	.00	0	.00	24	[illegible].64
		S	8	4.40	0	.00	11	1.21
		SS	5	2.75	0	.00	0	.00
		SC	1	.55	0	.00	3	.33
		SCI	0	.00	0	.00	4	.44
		X	0	.00	0	.00	3	.33
		CE	0	.00	0	.00	2	.22
		SCH	1	.55	0	.00	1	.11
/T/	7793		1361		184		2211	
		T	1329	97.65	183	99.46	2210	99.95
		TT	27	1.98	0	.00	1	.05
		ED	0	.00	0	.00	0	.00
		BT	0	.00	0	.00	0	.00
		TW	3	.22	0	.00	0	.00
		PT	1	.07	0	.00	0	.00
		CT	0	.00	0	.00	0	.00
		CHT	0	.00	1	.54	0	.00
		TH	1	.07	0	.00	0	.00
/T1/	411		141		15		70	
		TH	141	100.00	15	100.00	70	100.00
/T2/	149		21		5		35	
		TH	21	100.00	5	100.00	35	100.00

TABLE 20--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
227		18		140		1100		176		1281	
220	96.92	17	94.44	136	97.14	574	52.18	134	76.14	884	69.01
5	2.20	1	5.56	4	2.86	239	21.73	20	11.36	261	20.37
2	.88	0	.00	0	.00	217	19.73	18	10.23	135	10.54
0	.00	0	.00	0	.00	33	3.00	4	2.27	1	.08
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	33	3.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	4	.36	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0		0		4		300		23		90	
0	.00	0	.00	0	.00	89	29.67	2	8.70	0	.00
0	.00	0	.00	4	100.00	85	28.33	19	82.61	90	100.00
0	.00	0	.00	0	.00	41	13.67	1	4.35	0	.00
0	.00	0	.00	0	.00	51	17.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	16	5.33	0	.00	0	.00
0	.00	0	.00	0	.00	5	1.67	0	.00	0	.00
0	.00	0	.00	0	.00	5	1.67	1	4.35	0	.00
0	.00	0	.00	0	.00	1	.33	0	.00	0	.00
0	.00	0	.00	0	.00	4	1.33	0	.00	0	.00
0	.00	0	.00	0	.00	2	.67	0	.00	0	.00
0	.00	0	.00	0	.00	1	.33	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
366		62		98		1680		206		1625	
366	100.00	62	100.00	98	100.00	1465	87.20	201	97.57	1614	99.32
0	.00	0	.00	0	.00	182	10.83	4	1.94	2	.12
0	.00	0	.00	0	.00	18	1.07	1	.49	9	.55
0	.00	0	.00	0	.00	11	.66	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	1	.06	0	.00	0	.00
0	.00	0	.00	0	.00	2	.12	0	.00	0	.00
0	.00	0	.00	0	.00	1	.06	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0		0		0		141		24		20	
0	.00	0	.00	0	.00	141	100.00	24	100.00	20	100.00
2		2		0		71		3		10	
2	100.00	2	100.00	0	.00	71	100.00	3	100.00	10	100.00

TABLE 20--Continued

Phoneme	Total Occurrence	Grapheme	Initial					
			Primary		Secondary		Unaccented	
			f	%	f	%	f	%
/V/	1492		483		36		319	
		V	483	100.00	36	100.00	319	100.00
		LV	0	.00	0	.00	0	.00
		F	0	.00	0	.00	0	.00
/W/	626		344		95		78	
		W	344	100.00	95	100.00	55	70.51
		U	0	.00	0	.00	22	28.21
		OU	0	.00	0	.00	1	1.28
/Y/	120		40		5		73	
		I	1	2.50	0	.00	63	86.30
		Y	39	97.50	5	100.00	9	12.33
		E	0	.00	0	.00	1	1.37
/Z/	995		190		7		113	
		S	89	46.84	1	14.29	78	69.03
		Z	49	25.79	2	28.57	33	29.20
		ES	0	.00	0	.00	0	.00
		X	36	18.95	4	57.14	2	1.77
		ZZ	1	.53	0	.00	0	.00
		SS	11	5.79	0	.00	0	.00
		SC	3	1.58	0	.00	0	.00
		CZ	1	.53	0	.00	0	.00
/ZH/	102		3		0		56	
		SI	0	.00	0	.00	35	62.50
		S	1	33.33	0	.00	18	32.14
		G	2	66.67	0	.00	1	1.79
		Z	0	.00	0	.00	1	1.79
		TI	0	.00	0	.00	1	1.79

---

[a] Due to rounding off of percentage tabulations for each graphemic option representing each phoneme, the total percentage of occurrence of phoneme-grapheme correspondences for each phoneme may range between 99.96% and 100.04%.

TABLE 20--Continued

Medial						Final					
Primary		Secondary		Unaccented		Primary		Secondary		Unaccented	
f	%	f	%	f	%	f	%	f	%	f	%
17		0		1		410		35		191	
15	88.24	0	.00	1	100.00	405	98.78	35	100.00	191	100.00
2	11.76	0	.00	0	.00	2	.49	0	.00	0	.00
0	.00	0	.00	0	.00	3	.73	0	.00	0	.00
95		1		13		0		0		0	
83	87.37	1	100.00	0	.00	0	.00	0	.00	0	.00
12	12.63	0	.00	13	100.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
2		0		0		0		0		0	
2	100.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
5		0		91		390		48		151	
5	100.00	0	.00	91	100.00	281	72.05	37	77.08	58	38.41
0	.00	0	.00	0	.00	50	12.82	8	16.67	87	57.62
0	.00	0	.00	0	.00	35	8.97	3	6.25	6	3.97
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	22	5.64	0	.00	0	.00
0	.00	0	.00	0	.00	2	.51	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00
0		0		0		40		1		2	
0	.00	0	.00	0	.00	14	35.00	1	100.00	0	.00
0	.00	0	.00	0	.00	15	37.50	0	.00	0	.00
0	.00	0	.00	0	.00	10	25.00	0	.00	2	100.00
0	.00	0	.00	0	.00	1	2.50	0	.00	0	.00
0	.00	0	.00	0	.00	0	.00	0	.00	0	.00

/ZH/: < SI > represents this phoneme 100% of the time in final position in secondary accented syllables, although it is a single occurrence.< G > represents this phoneme 100% of the time in final position in unaccented syllables.

In summary, an analysis of the distribution of graphemic options for a given phoneme in each position in stressed and unstressed syllables clearly demonstrates that nearly all phonemes have at least one graphemic option exceeding the 80% criterion. It is interesting to observe that syllabic stress has a greater influence on the consonant phonemes than on the vowel phonemes. It will be recalled that when only positional effects were analyzed in terms of the distribution of graphemic options for a given phoneme, 21 exceeded the 80% criterion. However, when syllabic stress is included as a factor, only two /N1/ and /Z/, do not exceed this criterion. However, /ZH/ also could well be said not to exceed the 80% criterion since the two options which, when stress factors are considered, raise /ZH/ above this criterion are found in only three words.

It is apparent that the majority of phonemes, regardless of the phonemic classification, have at least one graphemic option exceeding the 80% criterion when the distribution of options for a given phoneme is analyzed in terms of position in syllables and in terms of position in stressed and unstressed syllables.

Some cautionary statements should be made, however. On the one hand, it is of interest to observe the degree to which the orthography can be seen to approximate the alphabetic principle when a thorough analysis is undertaken. It is another matter, however, to utilise this information without considering the frequency with which these graphemic options occur in the corpus. There is much possible value in the observation that the vowel phoneme /I3/ is represented by the graphemic option < I > in initial position in syllables nearly 90% of the time with a total of 1,535 occurrences. On the other hand, the utility of the observation that /O2/ is represented by < AU > nearly 94% of the time in final' position in unaccented syllables with a total of 15 occurrences is somewhat questionable. The graphemic option < I > representing /I3/ in that position can logically be called an approximation to the alphabetic principle. The graphemic option < AU > representing /O2/ cannot be said to enjoy that distinction. After all, a single occurrence of a particular phoneme-grapheme correspondence in some position may be said to exceed the 80% criterion, but if the phoneme occurs many times in different words in the lexicon a single occurring graphemic representation of that phoneme is a rare spelling and should be regarded as such. Such is the case for < SI > and < G > as representations of /ZH/ in final position in syllables. The three occurrences of these options in the corpus are rare spellings even though they are the only options used in that position in secondary accented and unaccented syllables.

The important generalization to be made from these data is this: When the phonological structure underlying the orthography is analyzed sequentially from simple phoneme-grapheme correspondences, through phoneme position in stressed and unstressed syllables, a hierarchical relationship is observed for both the phonological structure and the orthography. This disclosure has an important bearing on the inferences that can be made regarding how closely the orthography approximates the alphabetic principle and is the topic of the following chapter of this report.

CHAPTER 5

SOME INFERENCES FROM PHASE I

Introduction

One of the major purposes of this research was to determine the extent to which the American-English orthography may be described in terms of the phonological structure of the spoken language. The data suggest that this description is possible, at least to the extent that many phonemes can be shown to approximate the alphabetic principle when compared to an 80% criterion, and particularly when phoneme-grapheme correspondences are tabulated by their occurrences in certain positions in stressed and unstressed syllables.

It will be recalled that an ideal alphabetic orthography is that which has one, and only one, graphemic representation for each of the phonemes of oral language, and it will be recalled as well that some languages have writing systems which come quite close to achieving this ideal. The American-English orthography, on the other hand, is not so consistent when compared to this ideal as a *total* orthography. Viewing the orthography in its entirety, however, obscures the fact that many phonemes *individually* approximate the alphabetic principle quite closely, especially when positional and stress factors are considered.

The data in Phase I of this research have shown that more than half of the consonant phonemes have particular spellings which occur 80% or more of the time in the lexicon, irrespective of environmental factors. In addition, many of the vowel phonemes have particular graphemic representations 80% or more of the time in the lexicon, irrespective of these factors, although vowels collectively are less consistent than are the consonant phonemes.

Those phoneme-grapheme correspondences which exceed an 80% criterion, and with some frequency of occurrence in the lexicon, present no particular spelling problems. But what of the remaining correspondences which apparently have little relationship to the phonological structure underlying the orthography? Can descriptions be made of them which will indicate that, in certain environments, many phoneme-grapheme correspondences also can be said to approximate the alphabetic principle? The answers to these questions apply directly to conceptions of how spelling curriculums might be constructed and how instructional processes might be developed to parallel this content.

One way of describing an alphabetic orthography is to define it as a set of rules enabling an individual to transcribe the phonemes of oral language into graphic symbols which, when placed in the proper sequence, will produce visual surrogates of spoken words. Absent in most alphabetic orthographies, of course, are visual representations of the suprasegmental phonemes (the pitch, stress, and juncture phonemes) which are necessary to oral communication. Because these types of phonemes are not usually visually represented, dictionary-makers invent graphic devices which enable the dictionary user to convert selected words into acceptable spoken forms. Thus, symbols indicative of the degrees of stress given to syllables in these words are provided in most dictionaries. Levels of pitch, on the other hand, are not presented in dictionaries since these phonemes apply to the utterances of oral communication rather than to the individual words which form these utterances.

What factors concerning the phonology of oral language, then, do dictionaries contain which are applicable to studies of the orthography? Several relevant factors are noticeable. First, there is a description of the phonemes themselves and the graphic symbols which represent them. In addition, with the proper sorting, it can be seen that phonemes and graphemes have certain sequentially arranged patterns in syllables and words. And, finally, the influences of stress patterns upon both the pronunciation and spellings of syllables and words can be observed. In short, dictionaries are more than collections of selected words and their meanings; much information pertaining to the orthography itself may be found in dictionaries when this information is properly categorized and analyzed.

A word of caution must be given, however, regarding the use of dictionaries for studies of the orthography. It is quite true that the act of spelling entails transcribing the phonemes of oral language into the graphemes of written language. But, except for such specialized situations as spelling tests, the purpose of spelling is to communicate in visual form the intentions of the writer. In so doing, many words are used repeatedly, words which may contain inconsistent phoneme-grapheme correspondences. Thus, on the basis of frequency of occurrence of words in written communication, the orthography may appear to be hopelessly disparate with the alphabetic principle.

However, a study of the orthography is not primarily a study of the frequency with which certain words occur in running speech, nor their occurrence in written communication. A good speller must be able to transcribe any word in his speaking vocabulary into writing. In short, studies of the orthography are more concerned with phoneme-grapheme correspondence in the numbers of different words which occur in the lexicon than with the number of times a particular word may occur in written compositions.

What cues are available in the orthography which enable an individual to spell (write) successfully any word in his speaking vocabulary? In an alphabetic orthography such as American-English, the phonological structure underlying it provides the richest sources of information. It can be supposed that a good speller intuitively makes associations between the structure of the orthography and the structure of the phonology underlying it and devises his own cues from such factors as positional restrictions upon phoneme-grapheme correspondences and the way in which certain stress factors impinge upon the spelling of some syllables and words.

In effect, the act of spelling is analogous to the kind of analysis undertaken in this research, an act which entails sorting from a reservoir of phoneme-grapheme correspondences correct sequences of graphemes which correspond to the sequences of phonemes of the spoken word. For those phonemes closely approximating the alphabetic principle underlying an ideal alphabetic orthography, no other cues necessarily are needed. This phase of the research has indicated which phonemes may be satisfactorily spelled at this level. But for phonemes less consistent with the alphabetic principle, a further searching for cues which are inherent in the phonological structure underlying the orthography becomes necessary.

Thus, one may turn to the positions in syllables and words in which inconsistent phoneme-grapheme correspondences occur and seek cues from certain positional restrictions that are placed upon these correspondences. This research has disclosed that the occurrences of certain phoneme-grapheme correspondences in certain positions in syllables does tend to restrict which graphemic option may represent a given phoneme. These cues, however, apply to fewer numbers of different words. In this sense, such cues to correct spelling are less productive; that is, they may be generalized only to words in the individual's vocabulary that contain uniform phoneme-grapheme correspondences that occur in the same positions in the syllables of these words.

However, should positional restrictions prove not to be especially useful in selecting the appropriate graphemic option for a particular phoneme, one may search for cues among the many ways in which syllabic stress influences phoneme-grapheme correspondences in words. But here, again, those cues which are revealed will apply to even fewer numbers of different words, since such phoneme-grapheme correspondences are restricted not only to certain positions in syllables but to syllables having particular stress patterns. In sum, each additional analysis of phoneme-grapheme correspondences serves to increase the consistency of these correspondences but in fewer numbers of words in the lexicon. Let us see how a comparison between the design of this research and the act of spelling applies particularly to the findings obtained from this investigation.

The mean percentage of phoneme-grapheme correspondences on a 52 phoneme classification is 73.13%, a figure which falls below the 80% criterion. The 30 consonant phonemes, as seen on page 79, on the other hand, are represented by 30 different graphemic options 83.99% of the time, a figure which approximates the alphabetic principle. The 22 vowel phonemes are represented by 22 different graphemic options only 62.27%

of the time, a figure falling somewhat short of the 80% criterion, hardly suggestive that the speller will make appropriate graphemic selections for this set of phonemes.

Note what happens, however, when an individual vowel phoneme is selected and analyzed in terms of how the phonological structure underlying the orthography tends to restrict which graphemic options may represent it in writing. /I3/, the sound heard in medial position in the word pin, and one of the most frequently used phonemes in the language, is represented by 22 different graphemic options in a 22 vowel phoneme classification, none of which exceeds the 80% criterion when phoneme-grapheme correspondences are analyzed irrespective of other phonological factors.

Of these 22 graphemic options, however, < I > and < Y > collectively represent /I3/ over 91% of the time. In effect, on the basis of the distribution of graphemic options for /I3/, nine times out of ten either of these two graphemic options will represent /I3/. But which of these two options should an individual choose in the majority of cases where this phoneme is used? An examination of their occurrences in various positions in syllables suggests a partial answer.

It is observed from the data in page 146, Table 17, that < I > occurs over 89% of the time in initial position in syllables, as in the word it, and over 82% of the time in medial position in syllables, as in the word pin, while in final position in syllables < Y > occurs approximately 60% of the time in the corpus, as in the word baby. Thus, on the basis of position in syllables, the speller may determine that < I > will represent /I3/ at least four times out of five in initial and medial positions in syllables and < Y > three times out of five in final position in syllables.

Are there other cues which may be helpful in bringing either of these graphemic options even nearer to a one-to-one correspondence with /I3/? Consider the effects of syllabic stress. As Table 19, page 162, indicates, when all the graphemic options for this phoneme are compared, < I > represents /I3/ over 98% of the time in initial position in primary accented syllables, over 86% of the time in initial position in unaccented syllables, and is the only graphemic option representing /I3/ in initial position in secondary accented syllables. < Y >, meanwhile, does not represent this particular sound at all in initial position of syllables, except in five rarely used words: larynx, onyx, pharynx, polyp, and sibyl. Furthermore, when final position is analyzed for the occurrences of < I > and < Y >, it is observed that these options occur in final position only in unaccented syllables.

But, are such increasing restrictions upon phoneme-grapheme correspondences and the cues which are obtained from these restrictions equally useful in spelling words? Percentage figures alone are misleading, since the frequency with which a given phoneme-grapheme correspondence occurs in different positions in stressed and unstressed syllables may vary considerably. /I3/, for example, occurs 7,815 times in the corpus and is represented by < I > 5,346 times and < Y > 1,801 times. Irrespective of positional and stress factors, were either of these options to attain the 80% criterion, an individual could apply a knowledge of their consistent representations of /I3/ to many words in the lexicon. Since neither graphemic option achieves an 80% criterion without a consideration of additional phonological factors, this assurance is not available to the speller.

Positional effects do raise < I > above an 80% criterion, however. But the data also indicate that this graphemic option now occurs 1,535 times in initial position and 2,772 times in medial position in syllables. Thus, the generalization that < I > most often represents /I3/ in either of these positions is increasingly accurate, but applies to fewer occurrences of this phoneme in the lexicon. When syllabic stress is considered, < I > nearly achieves a one-to-one correspondence with /I3/ in initial position in primary accented syllables, but only 192 times in the corpus, while its one-to-one correspondence with /I3/ in initial position in secondary accented syllables is accountable in only 200 occurrences in the corpus. It can, therefore, readily be seen that analyzing phoneme-grapheme correspondences in terms of succeeding levels of the phonological structure increases the approximation of certain phonemes to the alphabetic principle but with decreasing productivity in terms of the number of word tokens in which this knowledge may be applied.

Let us summarize the findings of Phase I of this study in terms of the phonological structure underlying the American-English orthography in order to illustrate the ways in which the orthography parallels the structure of the spoken language. Table 21 includes a summary of mean percentages of phoneme-grapheme correspondences and the number of different phonemes used in terms of each phonological factor in comparison to approximations to the alphabetic principle in each of three phonological categories: (1) the total occurrence of phoneme-grapheme correspondences in the corpus, (2) the total occurrence of these correspondences in initial, medial, and final positions in syllables, and (3) the total occurrences of these correspondences in each position in stressed and unstressed syllables.

TABLE 21

PERCENTAGES OF APPROXIMATION TO THE ALPHABETIC PRINCIPLE OF PHONEME CLASSIFICATIONS AT EACH LEVEL OF THE PHONOLOGICAL STRUCTURE UNDERLYING THE ORTHOGRAPHY

Phonological Level	52 Phon. Class		22 Vowel Phon.		30 Cons. Phon.	
	No. Phon. Used	%	No. Phon. Used	%	No. Phon. Used	%
Phoneme-Grapheme Correspondences	52	73.13	22	62.27	30	83.99
Position Factors						
Initial	45	81.61	20	76.93	25	86.28
Medial	45	78.06	22	66.11	23	96.47
Final	43	79.78	18	80.90	25	78.67
Mean %		79.82		74.65		87.14
Position & Stress						
Initial						
Primary	44	83.73	19	77.23	25	90.22
Secondary	35	89.17	11	85.53	24	92.80
Unaccented	42	84.92	17	79.78	25	90.05
Mean %		85.94		80.85		91.02
Medial						
Primary	42	81.86	21	66.89	21	96.82
Secondary	33	83.04	20	71.12	13	94.96
Unaccented	39	89.58	22	80.74	17	98.41
Mean %		84.83		72.92		96.73
Final						
Primary	36	77.12	13	79.01	23	75.23
Secondary	33	81.67	11	80.75	22	82.58
Unaccented	43	85.45	18	82.42	25	88.47
Mean %		81.69		80.73		82.09
Average of Mean %		84.15		78.17		89.95

Phoneme-Grapheme Correspondences

Table 21 shows that 22 vowel phonemes, in a 52 phoneme classification, are represented by 22 graphemic options 62.27% of the time, although none of the phonemes in this category has only one graphemic option representing it. /05/ does have two graphemic options only, < 0 > and < 0-E >, but both /I3 and /ə/ are represented by 22 different graphemic options when all occurrences of these phonemes are tabulated, with no single graphemic option achieving an 80% criterion, irrespective of other phonological factors.

Thirty consonant phonemes are represented by 30 graphemic options 83.99% of the time when all occurrences of these phonemes are tabulated, with /HW/, /M1/, /T1/, and /T2/ each having one graphemic option representing it in the orthography. /SH/, on the other hand, has 15 different graphemic options representing it, none of which exceeds the 80% criterion when all occurrences of this phoneme are tabulated.

Collectively, all phonemes in a 52 phoneme classification are represented by 52 graphemic options 73.13% of the time for all occurrences in the corpus.

Just as there is a wide range in the number of graphemic options which may represent each phoneme, there is also a considerable variation in the number of times each phoneme occurs in the different words of the corpus. For example, in a 22 vowel phoneme classification, /05/ occurs 127 times and /I3/ occurs 7,815 times, while /HW/ occurs 89 times and /R/ occurs 9,390 times among all occurrences of the 30 consonant phonemes.

Thus, it can be seen that the degree to which a given phoneme may be said to approximate the alphabetic principle is in part dependent upon the frequency of occurrence with which it is represented by a particular graphemic option ~~option~~ in all the words of the corpus. The utility of these data for the speller, however, is dependent upon the frequency with which a phoneme occurs in his speaking (and writing) vocabulary. Obviously, a phoneme which occurs frequently in many different words, and in close correspondence with the alphabetic principle, presents less of a spelling problem than a phoneme which also occurs frequently but with little correspondence to this principle or one which has a very uniform spelling but occurs infrequently in the lexison.

Irrespective of frequency, however, it is apparent that a number of phonemes can be correctly represented in writing simply on the basis of close correspondence with particular graphemic options, wherever these correspondences occur in words. Indeed, as a group, 19 of 30 consonant phonemes are represented by 19 different graphemic options, 63%, over 80% of the time in the corpus. Of the 22 vowel phonemes, only six, 27%, are represented by six different graphemic options above the 80% criterion for all of their occurrences in the corpus.

The Effects of Position in Syllables on Phoneme-Grapheme Correspondences

The wide discrepancies with which individual phonemes approximate the alphabetic principle, as well as vowel and consonant phoneme categories collectively, raises the question of whether their occurrence in various positions in syllables provides fruitful cues to their most common graphemic representations.

Table 21 provides a summary of the numbers of phonemes in each phonemic classification which occur in each position in syllables, as well as the mean percentage by which the phonemes occurring in each position are represented by an equal number of different graphemic options.

It can be seen that every phoneme in both vowel and consonant phoneme classifications does not occur equally in every position in syllables. In turn, the phonemes which legitimately do occur in certain positions are unequally distributed to the extent that a particular phoneme, and particular phoneme-grapheme correspondences, may be used predominantly in some position. To illustrate the effects of these positional restrictions, the data indicate that only 18 of the 22 vowel phonemes occur in final position in syllables and that these 18 are represented by particular graphemic options 80.90% of the time in this position.

The effects of positional restrictions upon both sets of vowel phonemes, upon consonant phonemes, and upon the two total phonemic classifications may be summarized as follows. When the 22 vowel phonemes, in a 52 phoneme classification, are analyzed without regard for positional effect, their mean percentage of correspondence to the alphabetic principle is 62.27%. When position is considered, the percentage increases to 74.65%. Likewise, when only six vowel phonemes in this classification exceeded the 80% criterion when positional restrictions were tabulated, 15 of these 22 phonemes, 68%, now exceed the 80% criterion.

Consonant phonemes show a concomitant increase in their approximations to the alphabetic principle, except in final position in syllables. In final position, consonant phonemes decline slightly in their approximations to this principle, showing 78.67% consistency compared to 83.99% when these factors were not accounted for. However, when the 30 consonant phonemes are tabulated for mean percentages of the occurrences of phoneme-grapheme correspondences in all positions, they also reveal how positional influences restrict the selection of graphemic options which legitimately represent them. (Occurrences of single consonant phonemes in medial position in syllables are included in these tabulations, rather than as parts of consonant phoneme clusters in initial and final positions in syllables as was earlier discussed. When 30 consonant phonemes are analyzed in this way, their mean percentage of correspondence to the alphabetic principle increases from 83.99% to 87.14%. The number of consonant phonemes which exceed the 80% criterion now increases from 19 to 21, a percentage increase from 63% to 70%.

When vowel and consonant phonemes are combined, and a similar analysis is made, the 52 phonemes now are represented by 52 different graphemic options 79.83% of the time, a close approximation to the 80% criterion.

Thus, when positional restraints upon phoneme-grapheme correspondences are analyzed, phonemic classifications now exceed or approach the 80% criterion which was selected as being a close approximation to the alphabetic principle. These consistent correspondences, however, will be applicable to fewer numbers of words.

The Effects of Syllabic Stress on Phoneme-Grapheme Correspondences

Since an accounting of positional restraints upon the selection of phoneme-grapheme correspondences discloses that the orthography is more nearly alphabetical when these environmental factors are considered, let us determine whether syllabic stress tends further to restrict these correspondences.

Table 21 also presents the number of different phonemes in each phonemic classification which occur in initial, medial, and final positions in syllables having three different amounts of stress. In each of these positions in stressed and unstressed syllables, the mean percentage of phoneme-grapheme correspondences in comparison to the alphabetic principle are given.

Comparing the percentage tabulations of the occurrences of vowel and consonant categories in particular positions in stressed and unstressed syllables with the percentages obtained for preceding levels of the phonological structure, it is observed that closer approximations to the alphabetic principle are attained. When the 22 vowel phonemes of a 52 phoneme classification are analyzed, as seen in Table 21, the mean percentage of their graphemic representations in the three positions in stressed and unstressed syllables is 78.17%. Furthermore, as seen in Table 19, 17 of these 22 phonemes, 77%, now exceed the 80% criterion.

The 30 consonant phonemes, in turn, are observed in Table 21 to be similarly influenced by syllabic stress. Also, as was seen in Table 20, 28 of the 30 consonant phonemes, 93%, now exceed the 80% criterion, an increase of seven phonemes over the 21 exceeding this criterion when only positional effects were considered. However, /ZH/, one of the seven phonemes, does so on the basis of two rarely used options < G > and < SI > in these environments. The mean percentage of their occurrences in all three positions in stressed and unstressed syllables is 89.95%.

When vowel and consonant phonemes are combined into a 52 phoneme classification, the effect upon their approximations to the alphabetic principle when both positional restraints and syllabic stress are considered is noticeable. As a group, these phonemes now approximate the alphabetic principle 84.15% of the time, as compared to 73.12% when neither positional nor stress factors are attributed to them.

Two rather interesting observations are noted in these analyses. A careful examination of the percentage tabulations in Table 21 reveals that, contrary to usual assumptions, the "short" vowel phonemes, as identified by occurrences in medial position in syllables, are the least consistently represented in the orthography when both positional restrictions and syllabic stress restraints are tabulated in terms of how they influence the selection of graphemic options for these phonemes. In addition, a further examination of these data indicates that phoneme-grapheme correspondences occurring in unaccented syllables are more consistent with the alphabetic principle, as defined in this study, than are the occurrences of these ocrrespondences in stressed syllables.

Quite possibly these conditions result from two related factors. First, it will be recalled that each subsequent phonological level tends to reduce the number of occurrences of phoneme-grapheme correspondences in the different words of this corpus. This reduction in the occurrence of all phonemes therefore reduces the number of opportunities in which highly frequent and inconsistently spelled phonemes will be tabulated as well.It should also be remembered that this research has been concerned with a tabulation of lexical material rather than with an analysis of the frequency with which certain words occur in running text, and that each kind of analysis requires different conceptions of what is meant by the word "frequency."

Most statements concerning the relative inconsistencies of phoneme-grapheme correspondences in American-English have been directed toward the frequency with which certain "long" vowel phonemes occur in certain words used quite often in running text. In this sense of the meaning of "frequency," numerous occurrences of certain inconsistently spelled "long" vowel phonemes and certain consistently spelled "short" vowel phonemes in frequently used words obscures the fact that "frequency," in terms of lexical material, is an indication of occurrences in different words, rather than in reoccurrences of certain words (Herdan, 1956). Obviously, the occurrences of certain phoneme-grapheme correspondences in syllables with varying degrees of stress must also be described in terms of the kind of frequency referred to: i.e., lexical frequency or frequency of word usage.

Second, the data indicate that some phonemes are eliminated entirely from certain positions in stressed and unstressed syllables. Accordingly, the phonemes which do occur in these environments, if they are phonemes with consistent spellings, will increase the degree to which phoneme-grapheme correspondences tabulated in these environments exceed the 80% criterion. At any rate, it is most likely that the ratio between highly used and inconsistently spelled phonemes and infrequently used but consistently spelled phonemes is altered, apparently in the direction of increasing approximations to the alphabetic principle.

In the final analysis, however, these data do illustrate the basic premise of this research: that the orthography of American-English closely parallels the phonological structure of the spoken language and that the orthography closely approximates the alphabetic principle as it has been defined in this research.

Internal Constraints

It has been observed that many phonemes and graphemes do not occur, or occur infrequently, in certain positions in syllables. These restrictions within a syllabic environment have been examined by a number of investigators (Gleason, 1961; Herdan, 1956; Miller, 1951; Spang-Hanssen, 1959; and Porter, 1961). But, in addition to certain positional restrictions, there is some evidence to support the notion that the order or sequence in which phonemes and graphemes occur in syllables and words is similarly constrained (Garvin, 1963; Higginbottom, 1962; and Osborne, 1954). In effect, the occurrence of a given phoneme-grapheme correspondence in a particular position in syllables is in part contingent upon which phoneme-grapheme correspondences precede and follow it (Garvin, 1963, p. 65).

The influence of internal constraints upon the orthography was not incorporated into the research design for Phase I of this study. Yet, it can be speculated that many occurrences of phoneme-grapheme correspondences in the corpus are restricted by this environmental factor. It is therefore quite possible that such internal constraints within syllables upon the selection of phoneme-grapheme correspondences in written American-English provides yet another characteristic of the phonological structure underlying the orthography. These relationships will be examined in Phase II.

Phase I has attempted to provide a detailed statistical description regarding the occurrences of phoneme-grapheme correspondences in each of several levels of the phonological structure underlying the orthography of American-English. It has been seen that these levels are hierarchically related, to the extent that each subsequent phonological level is contingent upon its predecessor. An examination of the orthography in terms of these relationships has provided a more refined description of how nearly the orthography attains the alphabetic principle at each of these levels than previously has been attempted.

However, are these relationships productive in the spelling of words? It is one thing to ascertain that some phonemes closely approximate the alphabetic principle; but it is another matter to determine how useful this knowledge is when such correspondences are combined with less consistent phonemic representations to reproduce words graphically. Obviously, a pupil engaged in the act of spelling must have absolute control over the selection of phoneme-grapheme correspondences throughout the words he reproduces in writing. Such problems as these await further testing.

Phase I has also indicated other aspects of the structure of both spoken and written American-English which are not primarily phonological in nature. These morphological, syntactic, and semantic factors quite possibly may provide additional cues to the manner in which many words in the lexicon are spelled.

CHAPTER 6

A DEFINITIONAL MODEL FOR THE ANALYSIS OF AMERICAN-ENGLISH ORTHOGRAPHY

The stated purpose of our research project was to examine the phoneme-grapheme relationships that exist in the orthography to determine basic cues for the improvement of spelling instruction in the elementary school.

As the research on Phase I progressed and the data emerged, it became apparent that there were decreasingly productive levels of analysis of factors affecting the phoneme-grapheme relationships. The basic three levels of analysis were not new; linguists have traditionally divided the study of language into phonology, morphology, and semantics or syntax. The productive insight gained from the Phase I data was that these three levels appeared to present a heuristic model for the analysis of the orthography of the American-English language. This point may at first seem too obvious: since the written language is a surrogate for the spoken language, a model for the structure of the latter should surely be useful for the structure of the former. As Paul Garvin points out (1963, p. 22):

> . . . . written English . . . . although it is far from being a complete isomorph of Spoken English--is as much of a natural language as any of the spoken dialects, since it exhibits all the structural properties set forth in the definitional model of language, although, of course, their substantive manifestation is graphic rather than phonetic.

There are two points to be noted relative to the relationship between spoken and written language and our use of the following model:

1. As Garvin said, the written language is not a complete isomorph of the spoken language. The difference is important.
2. Our use of the model is not to analyze the structure of the written language qua language, but the rules for writing the written language on the basis of the spoken.

Let us first consider the possible sources of a written language. One could hypothesize a written language which was completely independent of any spoken language. In such a language, phonology, of course, could have no place whatsoever.

Secondly, there might be a written language and a spoken language which existed in such a relationship that the written language was basic to the spoken language. This is difficult to imagine, but say, for example, that an isolated community of deaf-mutes had existed since primeval time and developed a written language. Later this tribe was discovered by modern man and their written language was "transcribed" into a spoken language. Awkward as this concept is to imagine, it is the contention of the research team that the elementary school language program of the past half-century or so has behaved as if this were in fact the case: that the written language is the basis for the spoken. It is some such assumption that underlies the teaching methodology broadly known as the "phonics approach" to reading, whereby letters have sounds. If our American-English language were, in fact, such a type as we have here described, our usage of the model we are developing would be inappropriate; in place of phoneme or phonology in the model, we would have grapheme or graphology.

Thirdly, we have what is the typical relationship between a spoken and written language: the spoken language is basic and original;

the written language is secondary and derived.

Another fundamental way of looking at the relationship between the spoken and written forms of a language and the way that our model may be applied to these is to consider the different kinds of written language: logographic, morphographic, syllabic, and alphabetic.[1]

Garvin (p. 5) has stated that a theoretical model

> should be capable of accounting for the two apparently contradictory conclusions that have emerged as basic insights from the experience of the profession. One is that all the languages of the world are similar to each other, in that they are all languages. The other is that all the languages of the world are dissimilar to each other, in that they are all different languages. Any theoretical framework or model must account for both the similarity and the dissimilarity in such a way that they become logically compatible.

Therefore, our model of the orthography should be applicable to orthography in general, in its relationship to spoken language in general, and it should also be capable of application to any particular orthography to indicate its particularity.

Thus, a logographic orthography would have an underlying phonology, but it would be primarily related to the morphology and syntax of the language. At the other end of the continuum, a purely alphabetic language would have its orthography determined primarily at the phonological level, with morphological and syntactical factors determining only a small percentage of graphemic options. Nevertheless, since in one sense the language as a means of communication cannot be thought of in units smaller than the sentence, the orthography can never be divorced completely from syntax or semantics.

---

[1] Most natural language systems are mixtures of these various types. For example, in English the use of numerals is logographic rather than alphabetical. Nevertheless a language may usually be categorized as basically one or the other of these systems. Also, the terms "logographic" and "morphographic" might well be subsumed under the cover term "ideographic" with the symbols representing meaningful units in the linguistic system. There is no absolute consensus among linguists as to the usage of these terms or the classification of all languages as belonging to one or the other; nevertheless the substance of the argument here presented does not depend upon a precise definition of these terms.

Our model, then, can serve for the commonalities of all othographies of natural language in that they all would be determined to some degree by all three primary levels: phonological, morphological, and syntactical. It can serve to differentiate particular orthographies to the extent that they rely for their determination, with varying degrees of emphasis, upon each of the three basic levels of the model, and, as we shall see, upon sub-levels within each of the three primary levels.

So far, then, we have shown that the relationships between spoken and written language is such that a definitional model based upon the former can be used for a descriptive analysis of the latter. Again, in Garvin's words (p. 22):

> This means that although descriptive linguists have heretofore concentrated the bulk of their efforts on the analysis of spoken languages, the techniques of descriptive linguistics are equally applicable to written languages.
>
> . . . .
>
> Since a language is viewed as consisting of a system of units of different defining characteristics and orders of complexity which are linked to each other--and defined--by the relations into which they enter, the basic objective of descriptive linguistics becomes one of defining these units and of stating these relations.

This, then, becomes the rationale for our use of the model which follows and subsequently for the development of an algorithm for the spelling of the American-English orthography--Phase II of this project. Our basic unit is the phoneme-grapheme relationship; the relations of these units are determined by an analysis according to our definitional model.

The findings of the first stages of this phoneme-grapheme analysis were summarized in Phase I. The basic units, phoneme-grapheme relationships, it was found, are determined in American-English primarily at the phonological level of our model. With decreasing productivity, the factors of position of the phoneme within a syllable and the degree of stress upon the syllable aid in the determination of the graphemic option for a given phoneme. Phase I suggested that one further sub-level of the phonological remains to be studied: "internal constraints" or "environmental factors"--in other words, the graphemic option for a given phoneme is to some extent determined by which phonemes occurred immediately previous to and/or immediately following a given phoneme.

Thus, we have the following summation in graphic form of the phonological level of the model:

Phoneme-Grapheme Relationships Determined By:

1. Phonological factors
    1.1 Position
    1.2 Stress
    1.3 Environmental factors

Up through level 1.2 in the model, Phase I indicates that the spelling of individual phonemes in the American-English language can be predicted at a level of some 84% accuracy. We know from simple experience and observation that there are limiting factors in the environment of many words which reduce the number of graphemic options for a given phoneme. For example, when following the /S/ phoneme, the phoneme /F/ is always spelled < PH > rather than < F > (as in sphere). We can hypothesize that an analysis of such factors in the 17,000 word lexicon used for the current research project would increase the level of predictability for individual phonemes by several percent.

This would leave approximately 10+% of the phoneme-grapheme relationships indeterminate at the phonological level. The second basic level of the model, the morphological, must now be examined to determine what morphological relationships exist in the language that influence the choice of graphemic options for a given phoneme.

Again, common sense and examination of Phase I error lists enable us to divide our morphological level into several sub-levels which offer promise for analysis of the lexicon.

2. Morphological factors
   - 2.1 Compounding
   - 2.2 Affixation
   - 2.3 Word families[2]

Compounding. An examination of the ways in which independent morphemes combine to form compound words will help to produce rules for the spelling of certain sounds in certain environments. For example, Phase I provides statistical evidence to support what has long been taught as a rule, namely, that the /A/ sound in final position in a word is usually spelled < AY >. However, this rule is obscured in the spelling of a word such as playground. The /A/ is now no longer word final; application of the phonological rules would produce the spelling plaground. Thus, the level of predictability for the correct spelling of the long /A/ sound can be increased by taking into consideration the morphological factor of compounding of words.

---

[2] The use of the term "word family" here and following is another example of a popular rather than a linguistic usage. The term is simply used to describe groups of words with the same spelling option whether they are arbitrary (e.g., weigh, eight, freight, etc.) or whether they have something historically in common as to source.

Affixation. Consider the words *acute* and *account*. Phonologically considered, the spelling patterns of these two words are identical: /ə. K U T+ and /A3. K OU N T+/. Phonological rules and relatively regular phoneme-grapheme relationships account for the spelling of all the sounds in both words except the /K/ sound. In one case, it is spelled with one < C >; in the other with two, < CC >. The difference between the two may, however, readily be accounted for by consideration of simple rules of affixation and assimilation. Again, the level of predictability for the correct spelling of a phoneme has been increased by taking into consideration morphological factors--in this case, rules of affixation.

Word families. Relatively superficial examination of Phase I word lists shows that for many of the lesser-used graphemic options for given phonemes, groups of words can be readily noted that have common etymology. These tend to be borrowings from foreign languages. Again, one example: the relatively rare spellings of the long /A/ sound with < E > is seen to occur in a small group of words of Romance origin: *alcalde*, *andante*, *elite*, *finale*, *impresario*, *legato*, *machete*, *melee*, and *regime*. Once again, the predictability of the spelling of a phoneme has been increased by taking into consideration certain morphological factors--in this case the common familial origin of the roots of these words. It might be objected that this is a rather unnecessary refinement of the model--that in the last analysis you can account for the spelling of any word on an etymological basis. This, of course is true; but we are here considering the resolution of the spelling problem in a small residue of words, the bulk of which have been determined at "higher" or more productive and simpler levels of phonology and morphology. Our interest is in economy, from a logical point of view.

Finally, we can see that from the sound-to-spelling approach to the orthography, there would remain after all the above steps of analysis had been completed, a small group of words, the spelling of which could not be determined on phonological or on obvious morphological grounds. These words are the homonyms or homophones. Their spelling can only be determined by the context within which they are used.

While, again, it might be argued that the spelling of any word can ultimately be explained by the etymology of the word, in most cases of the homonyms this is not an economical method. Take the homonyms bear and bare. In Old English, the former is beran; the latter baer--a singularly unhelpful bit of information when confronted with the problem of spelling correctly the phonemes /B A2 R/ in a sentence. Thus, for practical purposes, the resolution of the spelling problem presented by homonyms lies in the area of syntax; one simply has to learn, by visual and haptical means, the different spellings for different uses in sentences.

Recapitulation. It may be helpful at this point to review the axioms and logical steps in the development of our model.

We begin with the assumption that in all modern natural languages, the orthography or the written language is based upon the spoken language.

The spoken language may be analyzed into phonological, morphological, and syntactical factors.

Combining these two assumptions, we can conclude that all orthographies will have phonological, morphological, and syntactical components. The influence of each of these in determining the structure of the orthography will depend upon the nature of the written form of the language (i.e., whether it is logographic, morphographic, syllabic, or alphabetic).

An alphabetical orthography is determined primarily by the phonological, to a lesser degree by morphological, and minimally by syntactical considerations.

The results of Phase I confirm the above assumption for American-English and suggest that the orthography is determined with decreasing productivity by the phonological factors of the unit phoneme-grapheme relationships as determined by position, stress, and environmental factors; by the morphological factors of compounding, affixation, and word families; and finally, in the case of homonyms, by syntactical factors.

Our completed definitional model now appears as follows: The orthography of American-English is determined by a set of rules for unit phoneme-grapheme relationships based, with decreasing productivity, upon three levels of analysis: phonological, morphological, and syntactical. The phonological level can be subdivided into considerations of position, stress, and environmental factors; the morphological level can be subdivided into considerations of compounding, affixation, and word families. This model may be summarized in tabular form thus:

Phoneme-Grapheme Relations Determined By:

1. Phonological factors
    - 1.1 Position
    - 1.2 Stress
    - 1.3 Environmental factors

2. Morphological factors
    2.1 Compounding
    2.2 Affixation
    2.3 Word families
3. Syntax